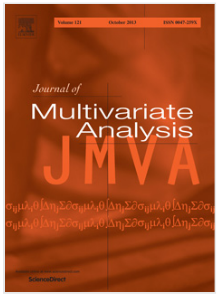
Accepted Manuscript



Linear transformations to symmetry

Nicola Loperfido

PII: S0047-259X(14)00101-8
DOI: http://dx.doi.org/10.1016/j.jmva.2014.04.018
Reference: YJMVA 3738

To appear in: *Journal of Multivariate Analysis*

Received date: 16 August 2013

Please cite this article as: N. Loperfido, Linear transformations to symmetry, *Journal of Multivariate Analysis* (2014), http://dx.doi.org/10.1016/j.jmva.2014.04.018

Linear Transformations to Symmetry

Nicola Loperfido
Dipartimento di Economia, Società e Politica
Università degli Studi di Urbino "Carlo Bo"
Via Saffi 42, 61029 Urbino (PU), ITALY
E-mail: nicola.loperfido@uniurb.it

Abstract

We obtain random vectors with null third-order cumulants by projecting the data onto appropriate subspaces. Statistical applications include, but are not limited to, the robustification of Hotelling's T^2 test against nonnormality. Our approach only requires the existence of the third-order moments and leads to normal transformed variables when the parent distribution belongs to well-known classes of sample selection models.

Some key words: Finite mixture; Multivariate analysis of variance; Nonrandom sampling; Singular value decomposition; Symmetrization.

1 Introduction

Let $\mu = (\mu_1, ..., \mu_d)^T$ be the mean of a $d-$dimensional random vector $x = (X_1, ..., X_d)^T$ satisfying $E(|X_i X_j X_k|) < +\infty$ for $i, j, k = 1, ..., d$. The third cumulant of x is the $d^2 \times d$ matrix $\kappa_3(x) = E\left\{(x-\mu) \otimes (x-\mu)^T \otimes (x-\mu)\right\}$, where "$\otimes$" denotes the Kronecker product (see, for example, De Luca & Loperfido, 2012). In the following, when referring to a third cumulant, we implicitly assume the existence of all the third-order moments of the corresponding random vector. The third cumulant of x is a null matrix when x is symmetric about a real vector $c \in \mathbb{R}^d$, that is if $x - c$ and $c - x$ are identically distributed. However, the converse is not necessarily true, as shown by many univariate examples. Bearing this in mind, we shall refer to random vectors whose third cumulants are null matrices as to weakly symmetric vectors. Weak symmetry, or lack of it, plays a fundamental role in probability and statistics. As a first example, the asymptotic distributions of commonly used MANOVA statistics greatly simplify when the sampled distribution is weakly symmetric (Gupta *et al*, 2008). As a second example, multivariate sample means admitting valid Edgeworth expansions converge to normality at a quicker rate, when the observations are weakly symmetric (Marsh, 2004). Similar comments hold for the asymptotic distribution of maximum likelihood estimates (Patriota & Cordeiro, 2011). As a third example, theoretical and empirical results (Mardia, 1970; Mardia, 1974; Everitt, 1979; Davis, 1982) hint that sampling distribution of Hotelling's T^2 statistic is quite robust to nonnormality, when the sampled distribution is weakly symmetric. Moreover, theoretical results in Fujikoshi (1997) imply that the Kolmogorov distance between the sampling distribution of Hotelling's statistic and the chi-squared distribution with d degrees of freedom converges to zero at a faster rate when the sampled distribution is $d-$dimensional, weakly symmetric, centered at the origin and has finite moments of appropriate order.

Symmetry is usually pursued by means of power transformations, primarily the Box-Cox one. Statistical applications include skewness removal from Hotelling's T^2 statistic when testing hypotheses about a multivariate mean (Freeman & Modarres, 2006; Niaki & Abbasi, 2007). However, power transformations suffer from some serious drawbacks, as pointed out by Hubert & Van der Veeken (2008) and Lin & Lin (2010), among others. In the first place, the transformed variables are neither affine invariant nor robust to outliers. In the second place, they might not be easily interpretable nor jointly normal.

For the sake of completeness, we shall mention two symmetrization techniques different from power transformations. Hall (1992) studied empirical transformations for removing most of the skewness of an asymmetric statistic using a monotone and an invertible cubic polynomial. Fujioka & Maesono (2000) also propose a transformation for removing skewness from U-statistics. Both transformations are limited to univariate data.

The present paper deal with the above issues by means of appropriate linear transformations, with special emphasis on Hotelling's T^2 statistic and nonrandom sampling. The approach is nonparametric in nature, since it applies to any multivariate data with finite third-order moments. It also leads to multivariate normal transformed variables, under some additional assumptions. Both real and simulated data encourage its use in statistical practice.

The rest of the paper is organized as follows. Sections 2 and 3 describe the symmetrization methods for the bivariate case and the multivariate case, respectively. Section 4 applies the method described in Section 3 to nonrandom samples from multivariate normal distributions. Sections 5 and 6 assess the practical relevance of the theoretical results in the previous sections by means of simulation studies and numerical examples, respectively. Section 7 contains some concluding remarks and hints for future research. All proofs are deferred to the Appendix.

2 The bivariate case

This section investigates the simplest case of linear transformations to symmetry, which involves two random variables only. We shall motivate it with the following example. Let Z_1, Z_2, Z_3 be three independent, identically distributed gamma variables. Also, let $W_1 = Z_1 - Z_3$ and $W_2 = Z_2 - Z_3$. Then W_1 and W_2 are symmetric random variables but the third cumulant of $w = (W_1, W_2)^T$ is not a null matrix. As a direct consequence, no componentwise power transformation $\left(W_1^a, W_2^b\right)^T$, with $a, b \in \mathbb{R}$, has a third cumulant which is a null matrix. However, there are three linear functions of w which are symmetric: W_1, W_2 and $W_1 - W_2$.

A natural question to ask is whether any two random variables with finite third moments X_1 and X_2 might be linearly combined to form another random variable $a_1 X_1 + a_2 X_2$ whose third cumulant is zero. Surprisingly enough, the answer is in the affirmative, as it can be shown constructively. When the third cumulant of either variable is zero, the linear function might be taken as the variable itself. Hence, without loss of generality, we shall assume that the third cumulants of both X_1 and X_2 are different from zero. The third cumulant $\kappa_3(W)$ of the random variable $W = wX_1 + X_2$ is $E\left\{(wY_1 + Y_2)^3\right\} = w^3 E\left(Y_1^3\right) + 3w^2 E\left(Y_1^2 Y_2\right) + 3wE\left(Y_1 Y_2^2\right) + E\left(Y_2^3\right)$, where $Y_1 = X_1 - E(X_1)$ and $Y_2 = X_2 - E(X_2)$. The third cumulant of W is then a cubic polynomial in w: $\kappa_3(W) = aw^3 + bw^2 + cw + d$, where $a = E\left(Y_1^3\right)$, $b = 3E\left(Y_1^2 Y_2\right)$, $c = 3E\left(Y_1 Y_2^2\right)$, $d = E\left(Y_2^3\right)$. By elementary algebra the the cubic equation $ax^3 + bx^2 + cx + d = 0$ has at least one real root, that is $s + t - v$, where $s = \sqrt[3]{r + u}$, $t = \sqrt[3]{r - u}$, $u = \sqrt{q^3 + r^2}$, $q = \left(3ac - b^2\right) / \left(9a^2\right)$, $r = \left(9abc - 27a^2 d - 2b^3\right) / \left(54a^3\right)$, $v = E\left(Y_1^2 Y_2\right) / E\left(Y_1^3\right)$.

Sample moments provide convenient choices for the variables X_1 and X_2 in many statistical applications. For example, let M_n and Q_n be the first and second sample moment of n random variables with finite sixth-order moments. Then there is a real value w such that the third cumulant of $wM_n + Q_n$ is zero. As a direct consequence, under very general conditions, it is possible to find an affine function of the first and second sample moment which converges to the standard normal distribution at a faster rate than the standardized sample mean itself.

The method naturally generalizes to any $d-$dimensional random vector $x = (X_1, ..., X_d)^T$, with $d > 2$ and finite third-order moments. Without loss of generality, we can assume that the variance of x is a positive definite matrix and that all components of x are standardized random variables whose third moments are different from zero. Also, let $\beta_1, ..., \beta_d$ be $d-$dimensional real vectors such that the $i-$th component of β_i is zero, for $i = 1, ..., d$. We can then apply the above described method to the pairs $\left(X_1, \beta_1^T x\right)$, ..., $\left(X_d, \beta_d^T x\right)$ to obtain the weakly symmetric random variables $Y_1 = \alpha_1 X_1 + \beta_1^T x$, ..., $Y_d = \alpha_d X_d + \beta_d^T x$, where $\alpha_i \in \mathbb{R}_0$, for $i = 1, ..., d$. Judiciously chosen vectors $\beta_1, ..., \beta_d$ will yield a vector $y = (Y_1, ..., Y_d)^T$ whose variance is a positive definite matrix. As an example, consider the trivariate random vector $x = (X_1, X_2, X_3)^T$ and apply the method described to the pairs (X_1, X_2), (X_2, X_3) and (X_3, X_1) to obtain the weakly symmetric random variables $Y_1 = \alpha_1 X_1 + X_2$, $Y_2 = \alpha_2 X_2 + X_3$, $Y_3 = \alpha_3 X_3 + X_1$. It follows that the variance of $(Y_1, Y_2, Y_3)^T$ is a positive definite matrix and that $\beta_1 = (0, 1, 0)^T$, $\beta_2 = (0, 0, 1)^T$, $\beta_3 = (1, 0, 0)^T$. However, the vector $y = (Y_1, ..., Y_d)^T$ is not in general weakly symmetric, despite the fact that all its components are. We shall deal with this problem in the next section, by making some assumptions about the rank of the third cumulant.

3 The multivariate case

This section deals with linear transformations to symmetry of several variables, motivated by the problem of testing the hypothesis that the mean $\mu = (\mu_1, ..., \mu_d)^T$ of a $d-$dimensional random vector $x = (X_1, ..., X_d)^T$ equals a known real vector $\mu_0 = (\mu_{01}, ..., \mu_{0d})^T$. In a nonparametric setting, the natural test statistic is Hotelling's T^2, whose distribution is approximately chi-squared when the sample size is large and the parent population is not too skewed. When this is not the case, the power method approaches the problem by looking for a transformation $y = \left(X_1^{\lambda_1}, ..., X_d^{\lambda_d}\right)^T$ which is symmetric, where $\lambda_i \in \mathbb{R}$ for $i = 1, ..., d$. Under the null hypothesis, the mean of y is in general different from $\left(\mu_{01}^{\lambda_1}, ..., \mu_{0d}^{\lambda_d}\right)^T$, and it is impossible to obtain without additional information about the distribution of x. On the other hand, assume the existence of a matrix $A \in \mathbb{R}^k \times \mathbb{R}^d$, with $1 \leq k < d$, such that $w = Ax$ is a weakly symmetric vector. Under the null hypothesis, the mean of w is known, being equal to $A\mu_0$. Weak symmetry of w improves the chi-squared approximation to the null distribution of Hotelling's T^2 statistics based on the transformed data. The following proposition shows that such a matrix exists when the third cumulant of x is not of full rank, and its rows belong to the linear space generated by the right singular vectors associated with its null singular values.

Proposition 1 *Let x be a $d-$dimensional random vector whose third cumulant $\kappa_3(x)$ has rank $d-k$, with $0 < k < d$. Moreover, let A be a $k \times d$ matrix whose rows span the null space of $\kappa_3^T(x)\kappa_3(x)$. Then the third cumulant of Ax is a null matrix.*

The above proposition might be applied to mixtures of multivariate normal distributions. Let the distribution of the $d-$dimensional random vector x be the mixture, with weights π_1, ..., π_k of the normal distributions $N_d(\mu_1, \Omega_1)$, ..., $N_d(\mu_k, \Omega_k)$. In the general case, the third cumulant of x is not a null matrix. Consider now a $k \times d$ matrix A satisfying $A\mu_i = 0_k$ for $i = 1$, ..., d. It follows that Ax is both centrally and weakly symmetric, being the scale mixture of normal distributions centered at the origin. Moreover, Ax is normal when the variances Ω_1, ..., Ω_k are equal.

Third cumulants of rank one provide the simplest, yet nontrivial applications of the above theoretical result. Statistical multivariate models whose third cumulants are rank one matrices include, but are not limited to, the extended skew-normal distribution (Arnold & Beaver, 2002), the normal-gamma distribution (Adcock & Shutes, 2012), a special case of the SGARCH model (De Luca & Loperfido, 2012) and mixtures of two homoscedastic symmetric components (Loperfido, 2013). Any third-order cumulant of rank one corresponding to a $d-$dimensional random vector might be represented as $\delta \otimes \delta^T \otimes \delta$, where δ is a $d-$dimensional, real, nonnull vector (see, for example, Christiansen & Loperfido, 2014). Hence weakly symmetric random vectors which are linear functions of a $d-$dimensional random vector x are of the form $\left(\alpha_1^T x, ..., \alpha_{d-1}^T x\right)$, where

$\alpha_i \in \mathbb{R}_0^d$ and $\alpha_i^T \delta = 0$, for $i = 1, ..., d-1$.

In practice, inference on the symmetrizing matrix A is sought. The matrix $\kappa_3^T(x)\kappa_3(x)$ might be estimated by $M_n = k_{3,n}^T(X)\,k_{3,n}(X)$, where $k_{3,n}(X)$ is the third sample cumulant $n^{-1}\sum_i (x_i - \overline{x}) \otimes (x_i - \overline{x})^T \otimes (x_i - \overline{x})$, where x_i^T is the $i-$th row of the data matrix X, $\overline{x}$ is the sample mean and n is the number of units. The eigenvalues and the eigenprojections of M_n converge to the corresponding quantities of $\kappa_3^T(x)\kappa_3(x)$, under mild assumptions (Tyler, 1981). This suggests to choose the rows of A among the right singular vectors associated with the smallest right singular values of $k_{3,n}(X)$. Also, asymptotic results in Tyler (1981) allow for hypothesis testing on A.

The symmetrization methods illustrated in the previous section and in this one complement each other. The former is applicable to any random vector of dimension greater than one, but it might lead to nonnegligible information loss. The former might retain more information, but it is not always applicable.

4 Weighted distributions

We shall now turn our attention to nonrandom samples, where the probability of including a given unit depends on the variables' outcome associated with the unit itself. Nonrandomness is usually modelled by weighted distributions of the form $f(x)\,w(x)/E\{w(x)\}$, where $f(x)$ is the sampled density, $w(x)$ is the weight function and $E\{w(x)\}$ is the normalizing constant (Patil & Rao, 1978). When the sampled distribution is multivariate normal and the weight function satisfies $0 \leq w(-a) = 1 - w(a) \leq 1$ for $a \in \mathbb{R}^d$ the weighted distribution is generalized skew-normal (Genton & Loperfido, 2005), with pdf $2\phi_d(x;\xi,\Omega)\,w(x-\xi)$, $\phi_d(x;\xi,\Omega)$ being the pdf of $N_d(\xi,\Omega)$. Several authors, including Genton & Loperfido (2005), Ley & Paindaveine (2010*a*, 2010*b*), Hallin & Ley (2012) also considered the special case $w(x-\xi) = h\left\{\alpha^T(x-\xi)\right\}$, where $h(\cdot)$ is a function satisfying $0 \leq h(-c) = 1 - h(c) \leq 1$ for any real value c and α is a $d-$dimensional real vector.

Let x_1, ..., x_n and y_1, ..., y_m be two samples drawn from $N_d(\mu_1,\Sigma)$ and $N_d(\mu_2,\Sigma)$, respectively, for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, where $\det(\Omega) > 0$ and $n + m > d$. The problem is straightforward under random sampling, but becomes very difficult when one or both samples are nonrandom. In the first place, the distribution of Hotelling's T^2 statistics is not known any more. Moreover, the two samples might be biased in different ways, so that $E(x_i) = \mu_1 + \zeta_1$ and $E(y_j) = \mu_2 + \zeta_2$, with $\zeta_1, \zeta_2 \in \mathbb{R}^d$, $i = 1, ..., n$ and $j = 1$, ..., m. As a direct consequence, the two sample means might be very similar while the populations means are not and vice versa. Similar problems arise in other multivariate techniques and cannot be solved by power transformations to symmetry.

The multivariate skew-normal distribution (Azzalini & Dalla Valle, 1996) is a very well-known model of weighted normal distribution, and we shall use it to evaluate the effect of biased sampling on the power of tests based on Hotelling's T^2 statistic. The distribution of a random vector x is multivariate skew-normal with location parameter ξ, scale parameter Ω and shape parameter α, that is $x \sim SN_d(\xi, \Omega, \alpha)$, if its pdf is $2\phi_d(x-\xi;\Omega)\,\Phi\left\{\alpha^T(x-\xi)\right\}$, where $\Phi(\cdot)$ is the cdf of a standardized normal variable and $\phi_d(x-\xi;\Omega)$ is the pdf of a $d-$dimensional normal distribution with mean ξ and variance Ω. Suppose we wish to use Hotelling's T^2 statistic for testing the hypothesis that the mean of a multivariate distribution is a null vector. Furthermore, assume that the sampled distribution is $N_d(0_d, \Omega)$, with $\det(\Omega) > 0$. Finally, assume that biased sampling causes the observations to be distributed as $SN_d(0_d, \Omega, \alpha)$, with $\alpha \neq 0_d$, so that they are skewed with means different from a null vector. The following theorem gives the asymptotic distribution of Hotelling's T^2 statistic when the above assumptions hold.

Proposition 2 *Let $\overline{x}$ and S be the mean and the variance of n observations randomly drawn from the distribution $SN_d(0_d, \Omega, \alpha)$, with $\det(\Omega) > 0$ and $\alpha \neq 0_d$. Then the asymptotic distribution of $\sqrt{n}\left(\overline{x}^T S^{-1}\overline{x} - 2\gamma\right)$ is normal with mean zero and variance $8\gamma\left\{1+\gamma+2(\pi-4)\gamma^2+4(\pi-3)\gamma^3\right\}$, where $\gamma = \alpha^T\Omega a/\left\{\pi+(\pi-2)\alpha^T\Omega a\right\}$.*

A statistician might then be willing to sacrifice a little information in order to perform well-known, reliable and simple statistical analyses. The following proposition might be helpful in pursuing such an approach.

Proposition 3 *Let x be a random vector whose pdf is $2\phi_d(x;\xi,\Omega)\,h\left\{\alpha^T(x-\xi)\right\}$, where $\phi_d(x;\xi,\Omega)$ is the pdf of $N_d(\xi,\Omega)$, $h(\cdot)$ is a function satisfying $0 \leq h(-c) = 1-h(c) \leq 1$ for any real value c and α is a $d-$dimensional real vector. Also, let A be a $(d-1)\times d$ matrix satisfying $A\Omega a = 0_{d-1}$. Then $y = Ax$ is a $(d-1)-$dimensional normal vector.*

The Closed Skew-Normal distribution introduced by Gonzalez-Farias *et al* (2003) provides another useful tool for modelling the sample bias. Its name reminds that it is closed with respect to conditioning, affine transformations and convolutions. Loperfido & Guttorp (2008) used it to model sample bias in air quality monitoring. The random vector x has a Closed Skew-Normal distribution of parameters $\xi, \Omega, \Psi, \nu, \Delta$, and we write $x \sim CSN(\xi,\Omega,\Psi,\nu,\Delta)$, if its density function is $\phi(x;\xi,\Omega)\;\Phi\left[\Psi(x-\xi);\upsilon,\Delta\right]/\,\Phi\left(0;\upsilon,\Delta+\Psi\Omega\Psi^T\right)$, where $x \in \mathbb{R}^p$, $\xi \in \mathbb{R}^p$, $\nu \in \mathbb{R}^q$, $\Psi \in \mathbb{R}^q \times \mathbb{R}^p$, $\Omega \in \mathbb{R}^p \times \mathbb{R}^p$, $\Delta \in \mathbb{R}^q \times \mathbb{R}^q$, $\Phi(z;\mu,\Sigma)$ is the cdf of $N_d(\mu,\Sigma)$ evaluated at $z \in \mathbb{R}^d$, while Ω and Δ are symmetric positive definite matrices. The following proposition gives a sufficient condition for the existence of a linear transformation Ax of x which is normal due to the removal of the sample bias.

Proposition 4 *Let $\kappa_3(x)$ be the third cumulant of the random vector x, where $x \sim CSN(\xi,\Omega,\Psi,\nu,\Delta)$, $\Omega \in \mathbb{R}^p \times \mathbb{R}^p$, $\Delta \in \mathbb{R}^q \times \mathbb{R}^q$ and $p > q$. Then the rank of $\kappa_3(x)$ is never greater than p nor q.*

5 A simulation study

This section uses simulations to assess the numerical performance of the proposed symmetrization method, when applied to a normal distribution contaminated by a small fraction of outliers. We simulated 10000 samples of size 100, 200, 300 and 400 from the normal mixture $\pi_1 N(0_d, I_d) + (1-\pi_1) N(10 \cdot 1_d, I_d)$, where $\pi_1 = 0.05$, 0.1, 0.15, and 0_d, 1_d, I_d are the $d-$dimensional vector of zeros, the $d-$dimensional vector of ones, the $d \times d$ identity matrix, respectively, for $d = 2$, 4, 6, 8, 10. Large sample size are needed since the sampling distributions of skewness and kurtosis are notoriously slow to converge to their limits, which are instrumental in assessing the normality of the transformed variables. Small values of π_1 exemplify situations where large values of skewness and kurtosis (Mardia, 1974) lead to sampling distributions of Hotelling's T^2 statistic which are very different from those obtained under the normality assumption (Davis, 1982).

Theoretical results in the previous section imply that a $d-$dimensional random vector might be linearly transformed into a $(d-1)-$dimensional normal random vector, when its distribution is as above. For each dimension and each weight, we computed the relative difference of the average value of Mardia's sample skewness from its expected value under normality. We also computed the relative frequency of samples for which the normality hypothesis was rejected at the 0.05 level when using the skewness-based test for multivariate normality proposed by Mardia (1970). Table 1 reports simulation results, which can be summarized as follows. Mardia's skewness suggests that the transformed variables are normal, consistently with the theoretical properties of the proposed method. Also, its performance does not seem to be noticeably influenced by the vector's dimension. The simulations' results are somewhat less satisfactory for samples of size 100, maybe due to the aforementioned slow convergence of Mardia's skewness.

		0.05		0.1		0.15	
100	**dim**	**err**	**rej**	**err**	**rej**	**err**	**rej**
	2	-7.68	4.46	-3.13	4.94	-11.51	3.78
	4	-4.63	4.54	-6.01	4.05	-8.46	3.94
	6	-4.56	4.12	-5.89	3.90	-8.44	3.02
	8	-3.84	4.20	-5.59	3.30	-8.21	2.06
	10	-3.63	3.81	-5.73	2.54	-8.49	1.27
200	**dim**	**err**	**rej**	**err**	**rej**	**err**	**rej**
	2	-1.65	4.70	-1.57	4.60	-1.08	4.90
	4	-1.56	5.20	-3.02	5.40	-4.08	4.10
	6	-1.23	4.60	-4.08	4.70	-3.42	4.90
	8	-2.12	4.30	-2.79	4.00	-3.34	4.10
	10	-2.03	3.60	-3.37	3.30	-3.97	2.70
300	**dim**	**err**	**rej**	**err**	**rej**	**err**	**rej**
	2	-2.18	4.69	-2.82	4.52	-4.01	4.57
	4	-1.77	4.75	-1.89	5.09	-2.28	4.86
	6	-1.85	4.79	-1.88	4.61	-2.28	4.59
	8	-1.64	4.36	-1.92	4.68	-2.54	3.92
	10	-1.45	4.50	-1.89	4.21	-2.24	3.83
400	**dim**	**err**	**rej**	**err**	**rej**	**err**	**rej**
	2	-1.37	4.68	-0.97	5.06	-4.15	4.57
	4	-0.82	5.17	-1.74	5.13	-1.72	5.10
	6	-1.21	5.18	-1.33	5.50	-1.87	4.62
	8	-0.77	5.16	-1.06	5.12	-1.85	4.28
	10	-1.08	4.34	-1.26	4.33	-1.82	4.18

Table 1: *err* denotes the relative difference between the average value of Mardia's sample skewness and its expected value under normality, multiplied by 100. Also, *rej* denotes the relative frequency of samples for which the normality hypothesis was rejected at the 0.05 level when using the skewness-based test for multivariate normality proposed by Mardia (1970), multiplied by 100.

6 A numerical example

This section uses the financial data analized by De Luca & Loperfido (2012) to illustrate the methods presented in the previous sections. They are percentage logarithmic daily returns (simply returns, henceforth) recorded from June 25, 2003 to June 23, 2008 in the financial markets of France, Netherlands and Spain (source: Morgan Stanley Capital International Inc.). Returns are arranged in the matrix X where the value in the $i-$th row and $j-$th column is the return in the $i-$th day and in the $j-$th market (markets being alphabetically ordered).

The third sample cumulant $k_{3,n}(X)$ is a 9×3 matrix with negative entries, consistently with previous theoretical and empirical results (De Luca & Loperfido, 2012). Its singular values are 2.130, 0.020 and 0.005, associated to the right singular vectors $v_1 = (0.587, 0.493, 0.642)^T$, $v_2 = (0.443, 0.467, -0.765)^T$ and $v_3 = (0.677, -0.734, -0.056)^T$. The first singular value is almost two hundred times greater than the sum of the remaining ones, suggesting that the symmetrization method based on Proposition 1 might be appropriate for the data at hand. The third standardized moments of the three European countries and those of the transformed variables are -0.349, -0.236, -0.461. The third standardized moments of the transformed variables Xv_2 and Xv_3 are much smaller, being -0.055 and -0.058.

Let Y denote the transformed data, that is the matrix whose columns are the transformed variables Xv_2 and Xv_3. We shall now compare the multivariate skewness of the original data X and the transformed data Y by means of several measures based on their third standardized cumulants. The measures of multivariate skewness proposed by Mardia (1970), Malkovich & Afifi (1973), Mori *et al* (1993) equal 0.256, 0.237, 0.256 for X and 0.014, 0.010, 0.016 for Y. Both univariate and multivariate measures of skewness clearly suggest that the proposed transformation was successful in removing virtually all of the skewness.

However, the practical relevance of these results might be questioned, since skewness is commonly believed to be less relevant than kurtosis in determining nonnormality of returns. We shall therefore examine the effect on kurtosis of the proposed symmetrization method. The fourth standardized cumulants of the three European returns are 4.657, 4.152 and 7.999. The fourth standardized cumulants of the transformed variables of Xv_2 and Xv_3 are much smaller, being 1.185 and 0.957. We shall now compare the multivariate kurtosis of the original data X and the transformed data Y by means of several measures based on their fourth standardized cumulants. The measures of multivariate kurtosis proposed by Mardia (1974), Malkovich & Afifi (1973), Mori *et al* (1993) equal 15.070, 66.427, 12.067 for X and 2.643, 1.425, 2.725 for Y. We conclude that, for the data at hand, the proposed transformation has a normalizing effect, since it significantly removes most of the skewness and excess kurtosis.

6.1 Conclusions

In this paper, we investigated two methods for removing skewness from data by means of linear transformations. Both methods overcome some limitations of well-known symmetrization methods. The first method applies to bivariate data and relies on the roots of third-degree polynomials. The second method relies upon the singular value decomposition and it is especially appropriate when sampling from multivariate, weighted distributions or normal mixtures. Both real and simulated data encourage their use in statistical practice.

We therefore recommend using the proposed methods whenever the performance of a statistical method is impaired by skewness, as it might happen when using Hotelling's T^2 statistic for testing the hypothesis that the mean of the sampled distribution is a null vector. We also recommend their use in conjunction with projection pursuit, a statistical technique aimed at finding interesting features of multivariate data by means of suitably chosen linear projections (see, for example, Huber, 1985). Once found, the interesting feature is removed in order to facilitate the search for other interesting features. Feature removal, also known as structure removal, constitutes then an important part of projection pursuit (Friedman, 1987). Projection pursuit might be based on skewness maximization (Huber, 1985), which has been successful in dealing with skew-normal distributions (Loperfido, 2010) and normal mixtures (Loperfido, 2013). Symmetrization procedures might then be used to remove skewness before looking for other features of the data, as for example kurtosis.

Both methods suffer from some limitations. The first method is better suited for bivariate data, while the second method might be unable to find the symmetrizing projections. Moreover, the transformed data suffer from some information loss, since the original data are projected onto a lower dimensional space. The loss might be substantial for the second method, when most of the singular values are positive. A researcher faced with this problem might be willing to stand a little skewness in order to retain more information. Subspaces spanned by singular vectors associated with small singular values are natural candidates for finding projected data with negligible skewness. We are currently investigating the problem.

Acknowledgments

The author wish to thank an associate editor and two anonymous referees for their insightful comments and suggestions that helped to improve the quality of the paper.

7 Appendix: proofs

Proof of Proposition 1. The matrix A might be decomposed into the product BC, where B is a nonsingular $k \times k$ matrix and C is a $k \times d$ orthonormal matrix whose rows span the null space of $\kappa_3^T(x)\kappa_3(x)$. The third cumulant of Cx is $\kappa_3(Cx) = (C \otimes C)\kappa_3(x)C^T$ (see, for example, De Luca & Loperfido, 2012). Consider now the identities $(C \otimes C)^T(C \otimes C) = \left(C^T \otimes C^T\right)(C \otimes C) = \left(C^T C \otimes C^T C\right) = I_{d^2}$, which follow from ordinary properties of the Kronecker product and from orthonormality of C. We can then write $\kappa_3^T(Cx)\kappa_3(Cx) = C\kappa_3^T(x)\kappa_3(x)C^T$. By construction, the rows of C span the null space of $\kappa_3^T(x)\kappa_3(x)$, so that $\kappa_3^T(Cx)\kappa_3(Cx)$ is a $k \times k$ null matrix. The third cumulant of Ax can be obtained from the third cumulant of Cx using the above mentioned property: $\kappa_3(Ax) = (B \otimes B)\kappa_3(Cx)B^T$. Since any product of a null matrix with another matrix is also a null matrix, the third cumulant of Ax is a null matrix and this complete the proof.

Proof of Proposition 2. Let μ and Σ be the mean and the variance of a random vector x whose distribution is $SN_d(0_d, \Omega, \alpha)$. Let $\delta = \Omega\alpha/\sqrt{1 + \alpha^T\Omega\alpha}$, so that $\alpha = \Omega^{-1}\delta/\sqrt{1 - \delta^T\Omega^{-1}\delta}$ and $\gamma = \delta^T\Omega^{-1}\delta/\left(\pi - 2\delta^T\Omega^{-1}\delta\right)$. The first, second, third and fourth cumulants of x are $\mu = \sqrt{2/\pi}\delta$, $\Sigma = \Omega - (2/\pi)\delta\delta^T$, $K_3 = \sqrt{2/\pi}(4/\pi - 1)\delta \otimes \delta^T \otimes \delta$, $K_4 = \left(8/\pi^2\right)(\pi - 3)\delta \otimes \delta^T \otimes \delta \otimes \delta^T$ (see, for example, De Luca & Loperfido, 2012; Loperfido, 2014).

The concentration matrix Σ^{-1} is the inverse of the sum of a matrix and a matrix product. We can then apply the formula $(A+BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1}+DA^{-1}B\right)^{-1}DA^{-1}$, where $A \in \mathbb{R}^k \times \mathbb{R}^k$, $B \in \mathbb{R}^k \times \mathbb{R}^m$, $C \in \mathbb{R}^m \times \mathbb{R}^n$, $D \in \mathbb{R}^n \times \mathbb{R}^k$ and all the necessary inverses exist (Mardia *et al*, 1979, page 459). By letting $A=\Omega$, $B=\delta$, $C=-2/\pi$, $D=\delta^T$ we obtain

$$\Sigma^{-1} = \Omega^{-1} - \frac{\Omega^{-1}\delta\delta^T\Omega^{-1}}{\delta^T\Omega^{-1}\delta-(\pi/2)} \text{ and } \Sigma^{-1}\delta = \frac{\pi\Omega^{-1}\delta}{\pi-2\delta^T\Omega^{-1}\delta}.$$

The asymptotic distribution of $\sqrt{n}\left(\overline{x}^TS^{-1}\overline{x}-\mu^T\Sigma^{-1}\mu\right)$ is normal with mean zero and variance $\lambda^T\Gamma\lambda$, where

$$\Gamma = \begin{pmatrix} \Sigma & K_3^T \\ K_3 & \Pi \end{pmatrix},\ \lambda = \begin{pmatrix} 2\eta \\ -\eta\otimes\eta \end{pmatrix},\ \eta = \Sigma^{-1}\mu = \frac{\sqrt{2\pi}\Omega^{-1}\delta}{\pi-2\delta^T\Omega^{-1}\delta}$$

and Π is the variance of $x\otimes x$ (Kollo & von Rosen, 2005, page 312). We shall now consider the summands which appear in the quadratic form

$$\lambda^T\Gamma\lambda = 4\eta^T\Sigma\eta - 4(\eta\otimes\eta)^T K_3\eta + \left(\eta^T\otimes\eta^T\right)\Pi(\eta\otimes\eta),$$

beginning with $4\eta^T\Sigma\eta = 4\mu^T\Sigma^{-1}\Sigma\Sigma^{-1}\mu = 4\mu^T\Sigma^{-1}\mu = 8\gamma$. Basic properties of the Kronecker product, together with the definitions of K_3 and η, lead to $-4(\eta\otimes\eta)^T K_3\eta = -4\sqrt{2/\pi}\,(4/\pi-1)\left(\eta^T\delta\right)^3 = 16(\pi-4)\gamma^3$.

The matrix Π is $E\left(y\otimes y^T\otimes y\otimes y^T\right) - vec(\Sigma)\,vec^T(\Sigma)$, where $y=x-\mu$ and $vec(\Sigma)$ is the vectorization of Σ (Kollo & von Rosen, 2005, page 285). The matrix K_4 is $E\left(y\otimes y^T\otimes y\otimes y^T\right) - (I_{d^2}+C_{d,d})(\Sigma\otimes\Sigma) - vec(\Sigma)\,vec^T(\Sigma)$, where $C_{d,d}$ is the $d^2\times d^2$ commutation matrix (Magnus & Neudecker, 1979; Loperfido, 2014). It follows that the quadratic form $\left(\eta^T\otimes\eta^T\right)\Pi(\eta\otimes\eta)$ equals $\left(\eta^T\otimes\eta^T\right)K_4(\eta\otimes\eta) + \left(\eta^T\otimes\eta^T\right)(I_{d^2}+C_{d,d})(\Sigma\otimes\Sigma)(\eta\otimes\eta)$. Basic properties of the Kronecker product, together with the definitions of K_4 and η, lead to $\left(\eta^T\otimes\eta^T\right)K_4(\eta\otimes\eta) = 32(\pi-3)\gamma^4$. By definition, $\eta=\Sigma^{-1}\mu$, so that $(\Sigma\otimes\Sigma)(\eta\otimes\eta) = (\Sigma\otimes\Sigma)\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)(\mu\otimes\mu) = \mu\otimes\mu$ and

$$\left(\eta^T\otimes\eta^T\right)(I_{d^2}+C_{d,d})(\Sigma\otimes\Sigma)(\eta\otimes\eta) = 4\gamma^2 + \left(\eta^T\otimes\eta^T\right)C_{d,d}(\mu\otimes\mu).$$

The identity $vec(M) = C_{d,d}vec(M)$ holds for any $d\times d$, symmetric matrix M (Magnus & Neudecker, 1979). Also, $\mu\mu^T$ is a $d\times d$, symmetric matrix and $\mu\otimes\mu = vec\left(\mu\mu^T\right)$. It follows that $\left(\eta^T\otimes\eta^T\right)C_{d,d}(\mu\otimes\mu) = \left(\eta^T\mu\right)^2 = 4\gamma^2$ and $\lambda^T\Gamma\lambda = 8\gamma\left\{1+\gamma+2(\pi-4)\gamma^2+4(\pi-3)\gamma^3\right\}$. This completes the proof.

Proof of Proposition 3. Without loss of generality we can assume that Ω is a positive definite matrix. Let $v_1,...,v_d$ be nonnull $d-$dimensional real vectors satisfying $v_1=\alpha$ and $v_i^T\Omega v_j = 0$ for $i\neq j$ and $i=1,...,d$. Also, let $y=Vx$, where $y=(Y_1,...,Y_d)^T$ and V is a $d\times d$ matrix whose $i-$th row is v_i^T for $i=1,...,d$. It follows that the components of y are mutually independent and the first component is univariate generalized skew-normal, while all other components are normally distributed. The inverse of V may be represented as $[A,b]$, where A is a $n \times (n-1)$ matrix and b is a $d-$dimensional real vector. Hence $x = Ay_{-1} + Y_1 b$, where $y_{-1} = (Y_2, ..., Y_d)^T$. Since y_{-1} and Y_1 are independent, the third cumulant of x is $\kappa_3(x) = \kappa_3(Ay_{-1}) + \kappa_3(Y_1 b)$, which might be simplified into $\kappa_3(x) = \kappa_3(Y_1 b)$ by recalling that Ay_{-1} is a normally distributed random vector. We shall now apply well-known properties of third cumulants of linear transformations to obtain $\kappa_3(x) = b^T \kappa_3(Y_1)(b \otimes b) = \kappa_3(Y_1)\left(b \otimes b^T \otimes b\right)$. The right singular vectors are the eigenvectors of $\kappa_3^T(x)\kappa_3(x) = \kappa_3^2(Y_1)\left(b^T b\right)^2 bb^T$. Hence there is only one positive singular value, and the corresponding right singular vector is proportional to b. By Proposition 1, the projection of x onto the subspace orthogonal to its dominant right eigenvector is a weakly symmetric random vector. The same subspace is spanned by the rows of A. Hence the weakly symmetric projection is also a $(d-1)-$dimensional normal random vector and this concludes the proof.

Proof of Proposition 4. Without loss of generality, we shall assume that $q < p$. Let y and w be two random vectors whose joint distribution is normal with $E(y) = \xi$, $E(w) = -\upsilon$, $var(y) = \Omega$, $var(w) = \Delta + \Psi\Omega\Psi^T$, $cov(y, w) = \Omega\Psi^T$. Also, let $A = \Omega\Psi^T\left(\Delta + \Psi\Omega\Psi^T\right)^{-1}$ and consider the decomposition $y = y - Aw + Aw$. Basic properties of normal random vectors imply that $y - Aw$ and w are independent, normal random vectors. Gonzalez-Farias *et al* (2003) showed that x and $y|w > 0$ are identically distributed, so that we can write $x \sim y - Aw + Aw_+$ where $w_+ = w|w > 0$. The third cumulant of the sum of independent random vectors is the sum of the third cumulants of the random vectors themselves, hence $\kappa_3(x) = \kappa_3(y - Aw) + \kappa_3(Aw_+)$. The identity $\kappa_3(x) = \kappa_3(Aw_+)$ follows from $y - Aw$ being a normally distributed random vector. Apply now multilinear properties of the third cumulant (see, for example, De Luca & Loperfido, 2012) to obtain $\kappa_3(x) = (A \otimes A)\kappa_3(w_+)A^T$. The rank of a matrix product is never greater than any of the matrices' ranks: $rank\{\kappa_3(x)\} \leq \min[rank\{\kappa_3(w_+), rank(A), rank(A \otimes A)\}]$. Also, the rank of the Kronecker product of two matrices is the product of the matrices' ranks, so that $rank\{\kappa_3(x)\} \leq \min\left[rank\left\{\kappa_3(w_+), rank(A), rank^2(A)\right\}\right]$. By assumption, $A \in \mathbb{R}^p \times \mathbb{R}^q$, $\kappa_3(w_+) \in \mathbb{R}^{q^2} \times \mathbb{R}^q$ and $q < p$. Hence we conclude that $rank\{\kappa_3(x)\} \leq q$.

References

Adcock, C.J. & Shutes, K. (2012). On the Multivariate Extended Skew-Normal, Normal-Exponential, and Normal-Gamma Distributions. *J. Statist. Theor. Prac.* **6**, 636-664.

Arnold, B. C. & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test* **11**, 7-54.

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

Christiansen, M.C. & Loperfido, N. (2014). Improved approximation of the sum of random vectors by the skew-normal distribution. To appear in *J. Appl. Prob.* **51**.

Davis, A.W. (1982). On the distribution of Hotelling's one-sample T^2 under moderate non-normality. *J. App. Prob.* **19**, 207-216.

De Luca, G. & Loperfido, N. (2012). Modelling multivariate skewness in financial returns: a SGARCH approach. *Eur. J. Fin.*, DOI:10.1080/1351847X.2011.640342.

Everitt, B.S. (1979). A Monte Carlo Investigation of the Robustness of Hotelling One- and Two-sample T^2 Test. *J. Amer. Statist. Assoc.* **74**, 48-51.

Freeman, J. & Modarres, R. (2006). Efficiency of $t-$Test and Hotelling's T^2-Test After Box-Cox Transformation. *Comm. Statist. - T. & M.* **35**, 1109-1122.

Friedman, J.H. (1987). Exploratory Projection Pursuit. *J. Amer. Statist. Assoc.* **82**, 249-266.

Fujikoshi, Y. (1997). An Asymptotic Expansion for the Distribution of Hotelling's T^2-Statistic under Nonnormality. *J. Multiv. Anal.* **61**, 187-193.

Fujioka, Y. & Maesono, Y. (2000). Higher order normalizing transformations of asymptotic U-statistics for removing bias, skewness and kurtosis. *J. Statist. Plann. Inf.* **83**, 47-74.

Genton, M. G. & Loperfido, N. (2005). Generalized Skew-Elliptical Distributions and their Quadratic Forms. *Ann. Inst. Stat. Math.*, **57**, 389-401.

Gonzalez-Farias, G. Dominguez-Molina, J.A., & Gupta, A.K. (2003). Additive properties of skew-normal random vectors. *J. Statist. Plann. Inf.* **126**, 521- 534.

Gupta A.K., Harrar S.W. & Fujikoshi Y. (2008). MANOVA for large hypothesis degrees of freedom under non-normality. *Test* **17**, 120–137.

Hall, P. (1992). On the Removal of Skewness by Transformation. *J. Roy. Statist. Soc. B* **54**, 221-228.

Hallin, M & Ley, C. (2012). Skew-symmetric distributions and Fisher information – a tale of two densities. *Bernoulli* **18**, 747-763.

Huber, P.J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13**, 435–475.

Hubert, M. & Van der Veeken, S. (2008). Outlier detection for skewed data. *J. Chemometrics* **22**, 235–246.

Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices.* Dordrecht: Springer.

Ley, C. & Paindaveine, D. (2010*a*). On the singularity of skew-symmetric models. *J. Multiv. Anal.* **101**, 1434–1444.

Ley, C. & Paindaveine, D. (2010*b*). On Fisher information matrices and profile log-likelihood functions in generalized skew-elliptical models. *METRON* **LXVIII**, 235-250.

Lin, T.C. & Lin, T.I. (2010). Supervised learning of multivariate skew normal mixture models with missing information. *Comput. Stat.* **25**, 183–201.

Loperfido, N. (2010). Canonical Transformations of Skew-Normal Variates. *TEST* **19**, 146-165.

Loperfido, N. (2013). Skewness and the Linear Discriminant Function. *Statist. Prob. Lett.* **83**, 93-99.

Loperfido, N. (2014). A note on the fourth cumulant of a finite mixture distribution. *J. Multiv. Anal.* **123**, 386–394.

Loperfido, N. & Guttorp, P. (2008). Network bias in air quality monitoring design. *Environmetrics* **19**, 661-671.

Magnus, J.R. & Neudecker, H. (1979). The commutation matrix: some properties and applications. *Ann. Statist.* **7**, 381-394.

Malkovich, J.F. & Afifi, A.A. (1973). On Tests for Multivariate Normality. *J. Amer. Statist. Assoc.* **68**, 176-179.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.

Mardia, K.V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B* **36**, 115-128.

Mardia, K.V., Kent, J.T. & Bibby, J.M. (1979). *Multivariate Analysis.* London: Academic Press.

Marsh, P. (2004). Transformations for multivariate statistics. *Econometric Theory* **20**, 963–987.

Mòri T.F., Rohatgi V.K. & Székely G.J. (1993). On multivariate skewness and kurtosis. *Th. Prob. Appl.* **38**, 547-551.

Niaki, S.T.A. & Abbasi, B. (2007). Skewness Reduction Approach in Multi-Attribute Process Monitoring. *Comm. Statist. T. & M.* **36**, 2313–2325.

Patil, G. P. & Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34**, 179-189.

Patriota, A.G. & Cordeiro G.M. (2011). A matrix formula for the skewness of maximum likelihood estimators. *Statist. & Prob. Lett.* **81**, 529–537.

Tyler, D.E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9**, 725-736.