JOURNAL OF MULTIVARIATE ANALYSIS 35, 203–222 (1990)

On the Cumulants of Affine Equivariant Estimators in Elliptical Families

RUDOLF GRÜBEL*

Faculty of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands

AND

DAVID M. ROCKE†

Graduate School of Management, University of California, Davis

Communicated by the Editors

Given a statistical model for data which take values in $\mathbf{R}^d$ and have elliptically distributed errors, and affine equivariant estimators $\hat{\mu}$ and $\hat{\Sigma}$ of a mean vector in $\mathbf{R}^d \otimes \mathbf{R}^n$ and a $d \times d$ scatter matrix, expressions are given for the covariances of the estimators in terms of their expectations and some unknown constants that depend on the model and the estimator. Higher order cumulants are also developed. These results place considerable constraints on the possible cumulants of $\hat{\mu}$ and $\hat{\Sigma}$, as well as those of estimators of higher order behavior such as multivariate skewness and kurtosis. These expressions are obtained using tensor methods.

1. INTRODUCTION

The purpose of this paper is to derive properties of estimators for statistical models with elliptical errors that depend only on the equivariance of the estimators. We find that the structure of the cumulants of such estimators is highly constrained and that many terms are necessarily zero and others are necessarily identical. Similar results are known for specific models and estimators, but our results have several advantages over previous work.

Received January 19, 1990.

AMS 1980 subject classifications: 62A05, 62E30, 62F11, 62F35, 62H10, 62J99.

Key words and phrases: maximum likelihood, multivariate location, multivariate regression, robust estimation, seemingly unrelated regression, tensor methods.

* Research supported by the Nuffield foundation.

† Research supported by the National Institute of Environmental and Health Sciences, National Institutes of Health Grant P42 ES04699.

0047-259X/90 $3.00

683/35/2-5

First, the results are quite general. We require of the estimator only that it is an equivariant function from the sample space to the space in which the statistic is defined. This includes linear estimators, maximum likelihood estimators, M-estimators, as well as the minimum covariance determinant estimator [22], the minimum volume ellipsoid estimator [21, 23] and related S-estimators [4], and Donoho's projection estimator [6, 7]. The model is required to have i.i.d. elliptical errors, but is otherwise general. We address the examples of multivariate location and scatter and multivariate regression, but the results apply equally to other linear and nonlinear models.

Second, the method developed here applies to any equivariant function. This means that one can develop information about estimators of the third and fourth cumulants of a distribution as well as information about the mean and variance. This should often ease the problem of higher order multivariate calculations. Also one may easily derive results about higher order cumulants of estimators of the mean and covariance matrix.

Third, in many cases, the results in this paper ease the problem of comparing estimators to each other or to asymptotic behavior. Some of the covariances in the theorems depend only on a few constants which are the same for spherical errors as for elliptical errors. For example, estimators of multivariate location and scatter have covariances that depend only on three constants. The higher order behavior is determined up to only a few more. This means that a small number of simulations under spherical errors can determine the constants and thus the behavior of the estimators in a particular case.

In the next section, we develop the tensor algebra machinery required to derive the key result that characterizes orthogonal invariant tensors. Section 3 presents the basic theorems that apply generally to equivariant estimators of models with elliptical errors. Section 4 presents applications of the basic result to several important specific cases such as multivariate location and scatter and multivariate regression. Section 5 summarizes the results.

2. Tensors and Tensor Spaces

2.1. *Basic Definitions*

If V and W are vector spaces over $\mathbf{R}$, then their tensor product $V \otimes W$ is a vector space over $\mathbf{R}$ generated as a vector space by elements $v \otimes w$ ($v \in V$, $w \in W$) subject to the bilinearity of the mapping $(v, w) \mapsto v \otimes w$. This can obviously be extended to a definition of the tensor product of any finite number of vector spaces which we represent as $V \otimes W \otimes \cdots$. In this paper, all vector spaces have finite dimension, although this is not required for the tensor product to exist (a more precise account of tensor products may be found in [20], for example). A typical element of this space is $\sum_{i=1}^{m} v_i \otimes w_i \otimes \cdots$. If the constituent spaces have bases $\{e_i\}$, $\{f_j\}$, etc., then the tensor product has basis $\{e_i \otimes f_j \otimes \cdots\}$. This means that a typical element is

$$\sum_{i, j, \ldots} t^{ij\cdots} e_i \otimes f_j \otimes \cdots = t^{ij\cdots}, \tag{2.1}$$

where the right-hand side is the common index notation convention of representing a tensor by writing a typical element (see [15], for example). This convention means that a symbol like t^{ijk} is ambiguously an entry in the tensor array and the entire array—the compactness of the notation is often worth the ambiguity.

To these three notations, the first drawn from the algebraic literature on tensor spaces and the last from the mathematical physics literature, we add sometimes a fourth—matrix notation—which is common in the statistical literature. If $V \cong \mathbf{R}^c$ and $W \cong \mathbf{R}^d$, then $V \otimes W$ can be identified with the space of $c \times d$ matrices by $v \otimes w \leftrightarrow vw^{\mathrm{T}}$, where v and w are identified with column vectors. The basis $e_i \otimes f_j$ of the tensor space is then identified with the basis e_{ij} for the space of matrices consisting of matrices all of whose elements are 0 except the (i, j) element, which is 1. A typical matrix A whose entries are $\{a^{ij}\}$ can be written as a tensor as $\sum a^{ij} e_i \otimes f_j$ or just a^{ij}. The ordinary action of a $c \times d$ matrix as a linear transformation $\mathbf{R}^d \mapsto \mathbf{R}^c$ is represented in the tensor space by extending the definition $(e_i \otimes f_j) \cdot f_k \equiv e_i f_j^{\mathrm{T}} f_k = e_i \delta_{jk}$, where δ_{jk} is Kronecker's δ, which is 1 if $j = k$ and 0 otherwise. This notation, however, fails for tensor products of more than two spaces. Thus matrix notation, which will suffice when the degree of the data is 1 or 2, is insufficient for more complex arrays.

If A is a linear transformation $V \mapsto V^*$, B is a linear transformation $W \mapsto W^*$, etc., then $A \otimes B \otimes \cdots$ is a transformation $V \otimes W \otimes \cdots \mapsto V^* \otimes W^* \otimes \cdots$ by $(A \otimes B \otimes \cdots)(v \otimes w \cdots) = Av \otimes Bw \otimes \cdots$. Of particular interest is when all of the constituent spaces are d-dimensional, so that the same linear transformation can act on all the spaces. In this case, if A: $V \mapsto V$ and if $v \otimes w \otimes \cdots \in V \otimes V \otimes \cdots$, then define

$$(v \otimes w \otimes \cdots)^A = Av \otimes Aw \otimes \cdots. \tag{2.2}$$

In index notation, if $s^{ij\cdots}$ is the result of the action of A on $t^{ij\cdots}$, then clearly

$$s^{ij\cdots} = \left(\sum_{i'j'\cdots} t^{i'j'\cdots} \right)^A = \sum_{i'j'\cdots} t^{i'j'\cdots} a_{ii'} a_{jj'} \cdots. \tag{2.3}$$

For tensors of degree 1 (vectors), this definition corresponds to the ordinary product of a matrix and a vector. For a tensor $T = t^{ij}$ of degree 2, the matrix equivalent of (2.3) is ATA^{T}. For tensors of higher degree, there is no equivalent matrix expression.

We will also need to extend this definition to spaces which consist of the p-fold tensor product V of $\mathbf{R}^d$ with another space W. In this case, if $u = v \otimes w$ with $v \in V$ and $w \in W$, and if A is a linear transformation on $\mathbf{R}^d$, then we define $u^A = v^A \otimes w$. This definition can obviously be ambiguous but the correct interpretation should usually be clear.

2.2. *Stochastic Tensors*

Let $\tilde{T}$ be a random variable taking values in a tensor product space $V \otimes W \otimes \cdots$. Then the expectation of $\tilde{T}$ is an element also of the tensor product space formed of the element-by-element expectations of the tensor array. If $\tilde{S}$ is another stochastic tensor with values in $V^* \otimes W^* \otimes \cdots$, then the covariance $\operatorname{Cov}(\tilde{T}, \tilde{S})$ is an element of

$$(V \otimes W \otimes \cdots) \otimes (V^* \otimes W^* \otimes \cdots) \cong V \otimes V^* \otimes W \otimes W^* \otimes \cdots \tag{2.4}$$

defined by

$$\operatorname{Cov}(\tilde{T}, \tilde{S}) = E(\tilde{T} \otimes \tilde{S}) - E(\tilde{T}) \otimes E(\tilde{S}). \tag{2.5}$$

If the two random variables take values in $\mathbf{R}^d$, then their covariance is a tensor of degree 2, i.e., a matrix. However, if $\tilde{T}$ and $\tilde{S}$ are themselves tensors of degree 2 or higher, then the covariance is a tensor with no ready interpretation as a matrix.

Some simple examples may clarify this point. Suppose $y_1, y_2, ..., y_n$ are random vectors in $\mathbf{R}^d$. Then each is a stochastic tensor of degree 1 and the collection can be considered a stochastic tensor of degree 2 which takes values in $\mathbf{R}^d \otimes \mathbf{R}^n$ (i.e., consider the data as a $d \times n$ matrix). The sample mean $\hat{\mu} = n^{-1} \sum_{i=1}^{n} y_i$ is also a stochastic tensor of degree 1 taking values in $\mathbf{R}^d$, so the covariance matrix of $\hat{\mu}$ is a tensor of degree 2 taking values in $\mathbf{R}^d \otimes \mathbf{R}^d$. The sample covariance matrix $\hat{\Sigma} = n^{-1} \sum_{i=1}^{n} (y_i - \hat{\mu})(y_i - \hat{\mu})^{\mathrm{T}}$ $\cong n^{-1} \sum_{i=1}^{n} (y_i - \hat{\mu}) \otimes (y_i - \hat{\mu})$ is also a tensor of degree 2. The covariances of the elements of $\hat{\Sigma}$ form a tensor of degree 4 taking values in $\mathbf{R}^d \otimes \mathbf{R}^d \otimes \mathbf{R}^d \otimes \mathbf{R}^d$. Often this covariance tensor of Σ is represented only laboriously by forming a vector of the nonredundant entries of $\hat{\Sigma}$. If the y_i vectors are not independent and identically distributed (i.i.d.), then one may need the covariance of the entire sample, which is a tensor of degree 4 in $(\mathbf{R}^d \otimes \mathbf{R}^n) \otimes (\mathbf{R}^d \otimes \mathbf{R}^n) \cong (\mathbf{R}^d \otimes \mathbf{R}^d) \otimes (\mathbf{R}^n \otimes \mathbf{R}^n)$ and thus can be represented as the sum of tensor products of $d \times d$ and $n \times n$ matrices, if desired.

Other cumulants of stochastic tensors can be calculated in the same way. If $\tilde{T}_i \in V_i$, then the rth moment is

$$E\left(\bigotimes_{i=1}^{r} \tilde{T}_i\right) \tag{2.6}$$

and the rth cumulant can be calculated using the formulae in [13 or 15].

2.3. *Orthogonal Invariant Tensors*

A $d \times d$ matrix $A = a_{ij}$ is *orthogonal* of $A^{T}A = I_d$, where I_d denotes the $d \times d$ identity matrix; equivalently, $\sum_{i=1}^{d} a_{ij}a_{ik} = \delta_{jk}$. Suppose $T = t^{ij\cdots}$ is a tensor in the p-fold tensor product of $\mathbf{R}^d$. Then T is *orthogonal invariant* if $T^A = T$, for all orthogonal transformations A. In index notation, this becomes

$$t^{ij\cdots} = \sum_{i'j'\cdots} t^{i'j'\cdots} a_{ii'} a_{jj'} \cdots. \tag{2.7}$$

Theorem 2.1 below characterizes orthogonal invariant tensors. One of the basis elements of the set of orthogonal invariant tensors of the tensor product of degree $p = 2q$ consists of the tensor τ_0 which has coefficient 1 for the basis elements $e_i \otimes e_i \otimes e_j \otimes e_j \otimes \cdots$ of the tensor product space and coefficient 0 for all other basis elements. Other generators of the space of orthogonal invariant tensors are those tensors satisfying a permutation of this condition. A few definitions are needed to make this precise.

Any basis element $e_{ij\ldots} = e_i \otimes e_j \otimes \cdots$ of the p-fold tensor product of $\mathbf{R}^d$ induces a partition $\pi(e_{ij\ldots})$ of the set of integers $\{1, 2, \ldots, p\}$ by placing those integers into the same class whose corresponding indices are the same. For example, $\pi(e_{1122}) = (12)(34)$ and $\pi(e_{1214}) = (13)(2)(4)$. If ω is a partition, we say that a basis element $e_{ij\ldots}$ is of *type* ω, if $\pi(e_{ij\ldots}) = \omega$ and is of *subtype* ω if $\pi(e_{ij\ldots}) \geqslant \omega$ in the lattice of partitions of $1, 2, \ldots, p$. For example, e_{112222} is of type (and subtype) $(12)(3456)$ and is of subtype $(12)(34)(56)$. If ω is a partition, define the tensors $\tau_\omega = \sum I(\pi(e_{ij\ldots}) = \omega)\, e_{ij\ldots}$ and $\tau^\omega = \sum I(\pi(e_{ij\ldots}) \geqslant \omega)\, e_{ij\ldots}$, where $I(\ )$ is the indicator function of the condition inside. We say ω is a *doubleton partition* if all classes consist of exactly two elements. Denote the set of all doubleton partitions of the integers $1, 2, \ldots, p$ by $\mathscr{D}_p$ and let $\bar{\mathscr{D}}_p = \{\omega \mid \exists \omega^* \in \mathscr{D}_p \text{ with } \omega \geqslant \omega^*\}$.

THEOREM 2.1. *The orthogonal invariant elements of the p-fold tensor product of* $\mathbf{R}^d$ *consist only of* 0 *if p is odd. If* $p = 2q$, *then the space of orthogonal invariant tensors has basis* $\{\tau^\omega \mid \omega \in \mathscr{D}_p\}$.

Proof. First, it is clear that the orthogonal invariant tensors form a subspace of the tensor product. Also each of the claimed basis elements is indeed orthogonal invariant. In order to avoid notational complexities, we will show this for the canonical doubleton partition $(12)(34)\cdots$ without loss of generality. We have

$$(\tau^{\omega})^A = \left(\sum e_{iijj\cdots}\right)^A$$
$$= \sum a_{i'i}a_{j'i}a_{k'k}a_{l'k}\cdots e_{i'j'\cdots}. \tag{2.8}$$

The sum over i yields $\delta_{i'j'}$, the sum over k yields $\delta_{k'l'}$, etc., so that the final sum is exactly τ^{ω}. This can, of course, be done for any partition by taking the summations by pairs. Note that this result remains true even if the dimension d is less than q, although partitions with more than d classes can obviously not occur.

To show that this exhausts the orthogonal invariant elements, consider first the orthogonal transformation defined by

$$\begin{aligned} e_i &\mapsto -e_i \\ e_j &\mapsto e_j, \qquad j \neq i. \end{aligned} \tag{2.9}$$

This shows that the index i must appear in an even number of places in every basis element with a nonzero coefficient, and this must be true for every index i. First, this means that no nonzero orthogonal invariant elements exist of an odd degree tensor product. Second, the tensor must be a linear combination of basis elements of subtype ω for $\omega \in \mathcal{D}_p$. Since any permutation of the basis elements of $\mathbf{R}^d$ is an orthogonal transformation, the coefficient of two basis elements of the same type must be the same.

Now suppose T is an orthogonal invariant tensor so that $T = \sum c_{\omega}\tau_{\omega}$, where the sum is over all $\omega \in \tilde{\mathcal{D}}_p$. The theorem asserts that there exists coefficients d_{ω} such that $T = \sum d_{\omega}\tau^{\omega}$, where the sum is over all $\omega \in \mathcal{D}_p$. Since $T^* = \sum c_{\omega}\tau^{\omega}$ is orthogonal invariant, so is $T - T^*$. We will therefore assume that $c_{\omega} = 0$ for all $\omega \in \mathcal{D}_p$ and show that this implies that $T = 0$. If not, then choose ω_0 such that $c_{\omega_0} \neq 0$ with the maximal number of classes in the partition. This means that $c_{\omega} = 0$ for all $\omega < \omega_0$. Also, there exists a class in ω_0 of $2l > 2$ elements, since the coefficients of all doubleton partitions were zero by hypothesis—without loss of generality, we may take this to be the first class in the partition.

Now consider the orthogonal transformation A defined on $\mathbf{R}^d$ by

$$\begin{aligned} A(e_1) &= e_1/\sqrt{2} + e_2/\sqrt{2} \\ A(e_2) &= e_1/\sqrt{2} - e_2/\sqrt{2} \\ A(e_j) &= e_j, \qquad j > 2. \end{aligned} \tag{2.10}$$

Consider also the element of type ω_0 which has the basis element e_1 in the first class, e_3 in the second class, e_4 in the third class, and so forth. This element has coefficient c_{ω_0} in T and coefficient $2^{-l}(2c_{\omega_0})$ in T^A, since it can arise only from itself and from the similar element with e_2 in the first class—all other elements which have this as a component of the image have zero coefficients by the maximality of ω_0. Since T is assumed to be orthogonal invariant, these two must be equal, which requires $c_{\omega_0}=0$, contrary to the choice of ω_0.

The above argument fails if $d<q$, since no doubleton partitions exist. The result is still true in this case, although the argument requires an additional step. If T is an orthogonal invariant element in this case, T is a linear combination of elements of type ω, where $\omega \in \tilde{\mathscr{D}}_p$. The minimal basis elements of T have d classes. If it could be shown that, by subtraction of a suitable element that is a linear combination of the τ^{ω}, for $\omega \in \mathscr{D}_p$, we could assume that all elements with d classes had zero coefficients; then the induction would proceed without difficulty. Thus it is necessary to show that, for any $T=\sum c_{\omega}\tau_{\omega}$, there exist coefficients d_{ω} such that $T-\sum d_{\omega}\tau^{\omega}$ has zero coefficients for all elements with d classes. Also, by considering the lattice embedding and the nature of the induction argument, we need only consider the case when $d=2$. The existence of such a set of coefficients follows from the linear independence of partitions and a simple count of the number of doubleton partitions versus the number of partitions with two classes. ∎

Remark. If V is the p-fold tensor product of $\mathbf{R}^d$ and $U=V\otimes W$ and if u is an orthogonal invariant element of U with respect to $\mathbf{R}^d$, then $u=\sum T_i\otimes e_i$, where each T_i is itself orthogonal invariant. If U has d-degree 2 and $W=\mathbf{R}^n\otimes\mathbf{R}^n$, then $U=I\otimes C$, for some $n\times n$ matrix C. Similar expressions can be worked out for mixed orthogonal invariant tensors of other d-degrees.

Remark. As noted in McCullagh's book [15], this result is known for dimensions up to 4 [25]; also see [11]. Also note that, if one requires only invariance to the special orthogonal group, in which the determinant is $+1$, the set of invariant tensors increases. The result for dimension 4 is given in [15].

3. Cumulant Tensors for Equivariant Estimators

In this section we derive a general result on affine equivariant estimators for statistical models with elliptically distributed errors. The class of models considered includes multivariate regression as well as multivariate location and scatter. The covariance tensors in some cases are determined up to only three unknown constants. We also consider higher order cumulants of $\hat{\mu}$ and $\hat{\Sigma}$ as well as the cumulants of affine equivariant estimators of multivariate skewness and kurtosis.

A distribution on $\mathbf{R}^d$ is spherically symmetric if its density f—we only consider absolutely continuous distributions—depends on its argument y only through the distance of y from the origin; i.e., if there exists a function $h: \mathbf{R} \mapsto \mathbf{R}$ such that $f(y) = h(y^{\mathrm{T}}y)$. Taking all nonsingular affine transformations of this distribution, we arrive at the d-dimensional elliptical family associated with h. If we fix h and the dimension d, then the family consists of distributions $Ell_{d,h}(\eta, \Omega)$, with density

$$f(y \mid \eta, \Omega) = |\Omega|^{-1/2} h((y-\eta)^{\mathrm{T}} \Omega^{-1}(y-\eta)), \tag{3.1}$$

where $\eta \in \mathbf{R}^d$ and Ω is a $d \times d$ positive definite symmetric matrix. Provided they exist, the mean of a random variable distributed as $Ell_{d,h}(\eta, \Omega)$ is η, and the covariance matrix is a multiple of Ω, which depends on h. These distributions are generalizations of the multivariate normal distribution in which $h(z) = (2\pi)^{-d/2} \exp(-z^2/2)$ and have been much studied [1–5, 12, 17–24, 26–28]).

Consider random vectors $y_i \in \mathbf{R}^d$. Given a sample $y = y_{si} \in \mathbf{R}^d \otimes \mathbf{R}^n$ of n values of y_i, let V be a subspace of $\mathbf{R}^d \otimes \mathbf{R}^n$ and suppose that $y = \eta + \varepsilon$, where $\eta \in V$, $\varepsilon \in \mathbf{R}^d \otimes \mathbf{R}^n$ and $\varepsilon_i \sim$ i.i.d. $Ell_{d,h}(0, \Omega)$. We will suppose V to be either fixed or randomly determined independent of ε. This setup includes many problems of multivariate analysis (where V is often $\mathbf{R}^d$ embedded diagonally in $\mathbf{R}^d \otimes \mathbf{R}^n$ by $v \mapsto \sum_{i,s} v_s e_s \otimes f_i$) as well as multivariate regression (where $V = \sum_s e_s \otimes \tilde{V}$, with $\tilde{V}$ a p-dimensional subspace of $\mathbf{R}^n$) and seemingly unrelated regression (where $V = \sum_s e_s \otimes \tilde{V}_s$, with $\tilde{V}_s$ a p_s-dimensional subspace of $\mathbf{R}^n$).

Suppose we have functions $\hat{\mu}: \mathbf{R}^d \otimes \mathbf{R}^n \mapsto V \subset \mathbf{R}^d \otimes \mathbf{R}^n$ and $\hat{\Sigma}: \mathbf{R}^d \otimes \mathbf{R}^n \mapsto \mathrm{PSD}(d) \subset \mathbf{R}^d \times \mathbf{R}^d$, where $\mathrm{PSD}(d)$ is the space of (symmetric) positive semidefinite $d \times d$ matrices. Suppose the functions are affine equivariant in the sense that, if A is a linear transformation on $\mathbf{R}^d$ and $v \in V$, then

$$\begin{aligned} \hat{\mu}(y+v) &= \hat{\mu}(y) + v \\ \hat{\mu}(y^A) &= \hat{\mu}(y)^A \in V^A \\ \hat{\Sigma}(y+v) &= \hat{\Sigma}(y) \\ \hat{\Sigma}(y^A) &= \hat{\Sigma}(y)^A. \end{aligned} \tag{3.2}$$

Now consider $z = \Omega^{-1/2}(y-\eta)$, where $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$, which is also a stochastic tensor in $\mathbf{R}^d \otimes \mathbf{R}^n$. Clearly, $z_i \sim$ i.i.d. $Ell_{d,h}(0, I)$. Also, if A is an orthogonal transformation, then Az_i has the same distribution as z_i so z^A has the same distribution as z and

$$\begin{aligned} E(\hat{\mu}_i(z))^A &= E(\hat{\mu}_i(z^A)) \\ &= E(\hat{\mu}_i(z)). \end{aligned} \tag{3.3}$$

Since the distribution of each $\hat{\mu}_i(z)$ is invariant under orthogonal transformations, its expectation is an orthogonal invariant tensor of degree 1 and, consequently, must be zero, so $\hat{\mu}$ provides an unbiased estimate of η. Also

$$\begin{aligned} E(\hat{\Sigma}(z))^A &= E(\hat{\Sigma}(z^A)) \\ &= E(\hat{\Sigma}(z)), \end{aligned} \tag{3.4}$$

so $E(\hat{\Sigma}(z)) = c_0 I$, for some scalar c_0, since it is an orthogonal invariant tensor of degree 2. Consequently,

$$E(\hat{\Sigma}(y)) = (c_0 I)^{\Omega^{1/2}} = c_0 \Omega. \tag{3.5}$$

Now the covariances of $\hat{\mu}(z)$ and $\hat{\Sigma}(z)$ can also be determined by the same methods. We have

$$\begin{aligned} E(\hat{\mu}_i(z) \otimes \hat{\mu}_j(z))^A &= E(\hat{\mu}_i(z^A) \otimes \hat{\mu}_j(z^A)) \\ &= E(\hat{\mu}_i(z) \otimes \hat{\mu}_j(z)) \\ E(\hat{\mu}_i(z) \otimes \hat{\Sigma}(z))^A &= E(\hat{\mu}_i(z^A) \otimes \hat{\Sigma}(z^A)) \\ &= E(\hat{\mu}_i(z) \otimes \hat{\Sigma}(z)) \\ E(\hat{\Sigma}(z) \otimes \hat{\Sigma}(z))^A &= E(\hat{\Sigma}(z^A) \otimes \hat{\Sigma}(z^A)) \\ &= E(\hat{\Sigma}(z) \otimes \hat{\Sigma}(z)). \end{aligned} \tag{3.6}$$

Thus, the covariance tensors are orthogonal invariant of degrees 2, 3, and 4, respectively and, transforming back to y-variates, we obtain the following theorem:

THEOREM 3.1. *Suppose $y, \varepsilon \in \mathbf{R}^d \otimes \mathbf{R}^n$ and $\eta \in V \subset \mathbf{R}^d \otimes \mathbf{R}^n$ are stochastic tensors with $y = \eta + \varepsilon$. Suppose that η is either fixed or random and independent of ε and suppose that $\varepsilon_i \sim$ i.i.d. $Ell_{d,h}(0, \Omega)$. Let $\hat{\mu}$: $\mathbf{R}^d \otimes \mathbf{R}^n \mapsto V \subset \mathbf{R}^d \otimes \mathbf{R}^n$ and $\hat{\Sigma}$: $\mathbf{R}^d \otimes \mathbf{R}^n \mapsto \mathrm{PSD}(d) \subset \mathbf{R}^d \otimes \mathbf{R}^d$ be arbitrary affine equivariant functions. Put $\Omega^{\otimes 2} = \Omega \otimes \Omega$ and let $\Omega^{\otimes 2*}$ and $\Omega^{\otimes 2**}$ be the tensors obtained from $\Omega^{\otimes 2}$ by the following index permutations; $(\Omega^{\otimes 2*})_{ijkl} = (\Omega^{\otimes 2})_{ikjl}$ and $(\Omega^{\otimes 2**})_{ijkl} = (\Omega^{\otimes 2})_{iljk}$. Then, whenever the expectations (which are with respect to ε, conditional on η) exist,*

$$E(\hat{\mu}(y)) = \eta \tag{3.7}$$

$$E(\hat{\Sigma}(y)) = c_0 \Omega \tag{3.8}$$

$$\mathrm{Cov}(\hat{\mu}(y), \hat{\mu}(y)) = \Omega \otimes C_1 \tag{3.9}$$

$$\mathrm{Cov}(\hat{\mu}(y), \hat{\Sigma}(y)) = 0 \tag{3.10}$$

$$\begin{aligned}\mathrm{Cov}(\hat{\Sigma}(y), \hat{\Sigma}(y)) &= c_2(\tau^{\omega_0})^{\Omega^{1/2}} + c_3(\tau^{\omega_1} + \tau^{\omega_2})^{\Omega^{1/2}} \\ &= c_2 \Omega^{\otimes 2} + c_3(\Omega^{\otimes 2*} + \Omega^{\otimes 2**}),\end{aligned} \tag{3.11}$$

where $\omega_0 = (12)(34)$, $\omega_1 = (13)(24)$, $\omega_2 = (14)(23)$, *and* C_1 *is a positive semidefinite* $n \times n$ *matrix such that* $\Omega \otimes C_1 \in V \otimes V$.

Theorem 3.1 asserts considerably more than what can be derived purely from the symmetry of $\hat{\Sigma}$ and of the covariance tensor. There are six distinct types of elements of the covariance tensor of $\hat{\Sigma}$ with itself, but only two undetermined constants in expression (3.11). Each of the four coefficients has a specific interpretation in terms of the original problem. First, c_0 is the multiplicative bias in the estimation of the covariance matrix; for example, in normal maximum likelihood, $c_0 = (n-1)/n$. Second, c_1^{ij} is the factor by which one multiplies the shape matrix of the elliptical distribution to obtain the covariance matrix for the mean estimators of observations i and j; for example, in normal maximum likelihood, this is $1/n$. Third, the covariance of $\hat{\sigma}_{ij}$ and $\hat{\sigma}_{kl}$ is $c_2 \sigma_{ij} \sigma_{kl} + c_3(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$. Note, however, that all of the constants depend on the sample size n, the dimension d, the estimators $\hat{\mu}$ and $\hat{\Sigma}$, and the function h of the elliptical distribution.

Under additional assumptions, the coefficients matrix C_1 can be further specified; in the next section we will show that C_1 reduces to a single constant in many cases, where V has a special structure. Here our next result deals with this matrix when $V = \sum_s e_s \otimes \tilde{V}$, as is the case in multivariate location and multivariate regression, and under the assumption of normal errors.

THEOREM 3.2. *Consider the situation of Theorem* 3.1.

(i) *Assume that* $V = \sum_s e_s \otimes \tilde{V}$, *where* $\tilde{V}$ *is a p-dimensional subspace of* $\mathbf{R}^n$ *and that the errors* ε_i *are i.i.d. multivariate normal with covariance matrix* Ω. *Let* $\hat{\mu}: \mathbf{R}^d \otimes \mathbf{R}^n \mapsto V$ *be an arbitrary affine equivariant onto function such that* $\hat{\mu}(y)$ *has finite covariance tensor* $\Omega \otimes C_1$. *Then there exists a* $p \times p$ *matrix* $\tilde{C}_1$ *which is positive semidefinite of rank* $\min(p, n-p)$ *and an orthogonal matrix* U *such that*

$$U C_1 U^{\mathrm{T}} = \begin{pmatrix} I + \tilde{C}_1 & 0_{p \times (n-p)} \\ 0_{(n-p) \times p} & 0_{(n-p) \times (n-p)} \end{pmatrix}.$$

(ii) *If* $\hat{\mu}\colon \mathbf{R}^d \otimes \mathbf{R}^n \mapsto V \subset \mathbf{R}^d \otimes \mathbf{R}^n$ *is linear and affine equivariant, then* $V = \sum_s e_s \otimes \tilde{V}$, *where* $\tilde{V}$ *is a p-dimensional subspace of* $\mathbf{R}^n$. *Furthermore, there exists a linear mapping* $\hat{\mu}\colon \mathbf{R}^n \mapsto \tilde{V}$ *so that* $\hat{\mu}(y) = \sum_{s=1}^d e_s \otimes \hat{\mu}(y_s)$.

(iii) *Suppose that the* ε_i *are iid multivariate normal with covariance matrix* Ω. *If* $\tilde{C}_1$ *is a* $p \times p$ *matrix which is positive semidefinite of rank* $\min(p, n-p)$, *and* $\tilde{V}$ *is an arbitrary p-dimensional subspace of* $\mathbf{R}^n$, *then there exists a linear affine equivariant function* $\hat{\mu}\colon \mathbf{R}^d \otimes \mathbf{R}^n \mapsto V = \sum_s e_s \otimes \tilde{V}$ *such that* $\Omega \otimes C_1$ *is its covariance tensor, where* C_1 *has the form given in* (i).

Proof. Clearly, it is sufficient to show this for $\Omega = I$ because of affine equivariance. Let $\tilde{P}$ be the orthogonal projection of $\mathbf{R}^n$ onto $\tilde{V}$ and P the induced projection of $\mathbf{R}^d \otimes \mathbf{R}^n$ onto V. Because $\hat{\mu}$ is affine equivariant we may write

$$\hat{\mu}(y) = \eta + P\varepsilon + \hat{\mu}((I-P)\,\varepsilon).$$

Now $P\varepsilon$ and $(I-P)\,\varepsilon$ are uncorrelated (since ε has identity covariance) and hence independent (since ε is normally distributed). Thus $P\varepsilon$ and $\hat{\mu}((I-P)\,\varepsilon)$ are uncorrelated. Suppose a basis of $\mathbf{R}^n$ is chosen in which the first p elements are a basis of V, with U being the orthogonal transformation to the new basis. Clearly, $P\varepsilon$ (considered as an element of $\mathbf{R}^d \otimes \mathbf{R}^p$) has identity covariance, and the covariance tensor of $\hat{\mu}((I-P)\,\varepsilon)$ must have the form (in the new basis)

$$\begin{pmatrix} \tilde{C}_1 & 0_{p \times (n-p)} \\ 0_{(n-p) \times p} & 0_{(n-p) \times (n-p)} \end{pmatrix}$$

so that the covariance $\hat{\mu}(y)$ is as given. The rank condition arises from dimensionality considerations.

For (ii), suppose that a linear estimator is defined by $\hat{\mu}_{si} = \sum_{t,j} P_{ij}^{st} y_{tj}$ and let a_t^u be a linear transformation on $\mathbf{R}^d$. Then the estimate of the transformed value of y is

$$\sum_{t,u,j} P_{ij}^{st} a_t^u y_{uj},$$

whereas the transformed value of the estimate is

$$\sum_t a_s^t \sum_{u,j} P_{ij}^{tu} y_{uj}.$$

Since y and a are arbitrary, we must have that $P^{st} = 0$ unless $s = t$ and that $P^{ss} = \tilde{P}$, for some $\tilde{P}$. This means that the sth component of $\hat{\mu}$ must be determined only from the sth component of y and using the same projection for each s.

To prove (iii), let $L: \mathbf{R}^n \mapsto \tilde{V}$ be the projection with matrix representation

$$\begin{pmatrix} I_p & B \\ 0_{(n-p)\times p} & 0_{(n-p)\times(n-p)} \end{pmatrix}$$

in the new basis in which the first p elements form a basis of V. Clearly, B can be chosen so that BB^{T} matches any desired matrix $\tilde{C}_1$ which is positive semidefinite of rank $\min(p, n-p)$. Define $\tilde{P}$ by $\tilde{P} = U^{\mathrm{T}}LU$ and define $\hat{\mu}$ by $\hat{\mu}_s(y) = \tilde{P}y_s$. Then this is a linear, affine equivariant estimator with the required covariance matrix. ∎

This theorem shows, under normal errors and the assumption about the structure of the subspace V that for each affine equivariant estimator $\hat{\mu}$ there exists a *linear* estimator with the same second-order structure. It will be interesting to examine a case where this result fails. This is provided by the discussion of seemingly unrelated regression (SUR) in the next section. Briefly, SUR has $V = \sum e_s \otimes V_s$, for possibly different subspaces V_s. For a given set of regressors, there exists no linear estimator (not depending on y) with the correct second-order structure.

Cumulants of higher order can easily be derived by the same method. We will illustrate this by developing the third and fourth cumulant tensors. Of the third cumulants, only $\mathrm{Cum}(\hat{\mu}, \hat{\mu}, \hat{\Sigma})$ and $\mathrm{Cum}(\hat{\Sigma}, \hat{\Sigma}, \hat{\Sigma})$ are nonzero, by considering the degree of the tensors involved. The third moment of $\hat{\mu}$, $\hat{\mu}$, and $\hat{\Sigma}$ is an orthogonal invariant tensor of degree four, and so must satisfy a similar equation to (3.11) with two other undetermined constants. Since the other terms in the expression for the third cumulant also are sums of similar terms, the third cumulant itself is of this form. The third cumulant tensor of $\hat{\Sigma}$ is of degree six, and thus is a sum of 15 terms, one for each of the doubleton partitions of six items. However, many of these must have the same coefficient by the symmetry of the cumulant operator and the symmetry of the $\hat{\Sigma}$. What remains are three terms, with therefore only three undetermined constants. One term consists of τ_0^{ω} for the canonical doubleton partition (12)(34)(56). The second consists of all τ^{ω}, where ω has one class in common with ω_0; for example, $\omega = (12)(35)(46)$. Otherwise put, these are the doubleton partitions whose partition lattice join with ω_0 is not the identity partition $1 = (123456)$. The third term consists of all τ^{ω}, where ω has no class in common with ω_0; for example, $\omega = (13)(24)(56)$. These are doubleton partitions whose partition lattice join with ω_0 is the identity partition (123456). This result, along with the similar result for the fourth cumulant tensors are given in the following theorem. We write $\kappa_{\mu\mu\mu}$ for $\kappa(\hat{\mu}(y), \hat{\mu}(y), \hat{\mu}(y))$, and similarly abbreviate the other cumulants.

THEOREM 3.3. *Under the assumptions of Theorem* 3.1, *the third cumulants of* $\hat{\mu}$ *and* $\hat{\Sigma}$ *are given by the following*:

$$\begin{aligned}
\kappa_{\mu\mu\mu} &= 0 \\
\kappa_{\mu\mu\Sigma} &= \Omega^{\otimes 2} \otimes C_4 + (\Omega^{\otimes 2*} + \Omega^{\otimes 2**}) \otimes C_5 \\
\kappa_{\mu\Sigma\Sigma} &= 0 \\
\kappa_{\Sigma\Sigma\Sigma} &= c_6 \Omega^{\otimes 3} + c_7 \sum_{\omega \vee \omega_0 < 1} (\tau^{\omega})^{\Omega^{1/2}} + c_8 \sum_{\omega \vee \omega_0 = 1} (\tau^{\omega})^{\Omega^{1/2}},
\end{aligned} \tag{3.12}$$

where $\omega_0 = (12)(34)$ *or* $(12)(34)(56)(78)$, *respectively*, $\omega_1 = (13)(24)$, $\omega_2 = (14)(23)$, *and the* C_i *are* $n \times n$ *matrices*; $\Omega^{\otimes 2}$, $\Omega^{\otimes 2*}$, *and* $\Omega^{\otimes 2**}$ *are defined in Theorem* 3.1 *and* $\Omega^{\otimes 3} = \Omega \otimes \Omega \otimes \Omega$. *The terms* $(\tau^{\omega})^{\Omega^{1/2}}$ *are tensors whose entries are permutations of the entries of* $\Omega^{\otimes 3}$. *The partition lattice join of* ω *and* ω_0 *is denoted* $\omega \vee \omega_0$.

The fourth cumulant tensors are given by the following, where $\nu(\omega)$ *denotes the number of classes of a partition* ω:

$$\begin{aligned}
\kappa_{\mu\mu\mu} &= \Omega^{\otimes 2} \otimes D_9 + (\Omega^{\otimes 2*} + \Omega^{\otimes 2**}) \otimes D_{10} \\
\kappa_{\mu\mu\mu\Sigma} &= 0 \\
\kappa_{\mu\mu\Sigma\Sigma} &= \Omega^{\otimes 3} \otimes C_{11} + \Bigg(\sum_{\omega \vee \omega_0 = \omega_1} (\tau^{\omega})^{\Omega^{1/2}} \Bigg) \otimes C_{12} \\
&\quad + \Bigg(\sum_{\omega \vee \omega_0 \in \{\omega_2, \omega_3\}} (\tau^{\omega})^{\Omega^{1/2}} \Bigg) \otimes C_{13} + \Bigg(\sum_{\omega \vee \omega_0 = 1} (\tau^{\omega})^{\Omega^{1/2}} \Bigg) \otimes C_{14} \\
\kappa_{\mu\Sigma\Sigma\Sigma} &= 0 \\
\kappa_{\Sigma\Sigma\Sigma\Sigma} &= c_{15} \Omega^{\otimes 4} + c_{16} \sum_{\nu(\omega \vee \omega_0) = 3} (\tau^{\omega})^{\Omega^{1/2}} \\
&\quad + c_{17} \sum_{\nu(\omega \vee \omega_0) = 2} (\tau^{\omega})^{\Omega^{1/2}} + c_{18} \sum_{\omega \vee \omega_0 = 1} (\tau^{\omega})^{\Omega^{1/2}},
\end{aligned} \tag{3.13}$$

where $\omega_1 = (12)(3456)$, $\omega_2 = (1234)(56)$, *and* $\omega_3 = (1256)(34)$ *and* D_9, $D_{10} \in \mathbf{R}^n \otimes \mathbf{R}^n \otimes \mathbf{R}^n \otimes \mathbf{R}^n$, *and the* C_i *are* $n \times n$ *matrices*; $\Omega^{\otimes 2}$, $\Omega^{\otimes 2*}$, $\Omega^{\otimes 2**}$, *and* $\Omega^{\otimes 3}$ *are as above, and* $\Omega^{\otimes 4} = \Omega \otimes \Omega \otimes \Omega \otimes \Omega$.

This formulation has several uses in investigating affine equivariant estimators. First, the conditions for the result to apply are easy to check. Then, many cumulant tensors require the calculation of only a few constants. Finally, these constants can be calculated for the spherical case $\Omega = I$ without loss of generality. Thus only a few integrations or a simple simulation can fix these constants. Once the constants have been calculated, comparison of different estimators becomes relatively routine.

For example, given simulated data for $k=1, 2, \ldots, N$, we can estimate c_2 and c_3 by considering the averages

$$S_2=\binom{d}{2}^{-1} N^{-1} \sum_{1\leqslant s<t\leqslant d} \sum_{k=1}^{N} (\hat{\sigma}_{ss}^k-\bar{\sigma}_{ss})(\hat{\sigma}_{tt}^k-\bar{\sigma}_{tt})$$
$$S_3=\binom{d}{2}^{-1} N^{-1} \sum_{1\leqslant s<t\leqslant d} \sum_{k=1}^{N} (\hat{\sigma}_{st}^k-\bar{\sigma}_{st})^2 \tag{3.14}$$
$$S_{23}=d^{-1}N^{-1} \sum_{1\leqslant s\leqslant d} \sum_{k=1}^{N} (\hat{\sigma}_{ss}^k-\bar{\sigma}_{ss})^2,$$

which estimate c_2, c_3, and c_2+2c_3, respectively. Linear combinations $u_2S_2+v_2S_3+w_2S_{23}$ and $u_3S_2+v_3S_3+w_3S_{23}$ estimate c_2 and c_3 if

$$\begin{aligned} u_2+w_2&=1\\ v_2+2w_2&=0\\ u_3+w_3&=0\\ v_3+2w_3&=1. \end{aligned} \tag{3.15}$$

Optimal weighting would depend on knowing the covariance tensor of the averages used, but equal weighting of each item is a reasonable alternative, which would lead to the estimates

$$\begin{aligned} \hat{c}_2&=\frac{d-1}{d+1}S_2-\frac{4}{d+1}S_3+\frac{2}{d+1}S_{23}\\ \hat{c}_3&=-\frac{4}{d+3}S_2+\frac{d-1}{d+3}S_3+\frac{4}{d+3}S_{23}. \end{aligned} \tag{3.16}$$

For large dimension d, use of S_2 and S_3 to estimate c_2 and c_3 would be essentially equivalent.

4. Applications

In this section we discuss several specific applications of the main result. First we consider the problem of multivariate location and scatter. Maximum likelihood for elliptical families provides a class of affine equivariant location estimators that are straightforward to described. Huber's M-estimators [9, 10, 14] provide a generalization, in which the location estimating equations and the scale estimating equations need not come as partial derivatives of the same likelihood. Then we discuss multivariate regression and a generalization, seemingly unrelated regression (SUR).

Although the general result described in the previous section is new, the application to the first two moments of estimators of multivariate location and shape have been previously described by many authors, in various forms in [8, 9, 16, 19, 24, 26–28]. Applications to robust estimation of location and shape are treated in [9, 26, 27, 28]. Note that none of these previous results has been extended to higher order cumulants of estimators of location and shape and that none has been extended to more general multivariate estimators such as multivariate regression.

4.1. *Affine Equivariant Estimators of Multivariate Location and Scatter*

In this case, the subspace V is the diagonal embedding of $\mathbf{R}^d$ in $\mathbf{R}^d \otimes \mathbf{R}^n$ and we may consider the function $\hat{\mu}$ to be into $\mathbf{R}^d$. Thus we have the following structure:

COROLLARY 4.1. *Suppose $y_i, \varepsilon_i \in \mathbf{R}^d$, $i = 1, ..., n$, and $\eta \in \mathbf{R}^d$ are stochastic tensors with $y_i = \eta + \varepsilon_i$. Suppose that η is either fixed or random and independent of ε and suppose that $\varepsilon_i \sim$ i.i.d. $Ell_{d,h}(0, \Omega)$. Let $\hat{\mu}$: $\mathbf{R}^d \otimes \mathbf{R}^n \mapsto \mathbf{R}^d$ and $\hat{\Sigma}$: $\mathbf{R}^d \otimes \mathbf{R}^n \mapsto \mathrm{PSD}(d) \subset \mathbf{R}^d \otimes \mathbf{R}^d$ be arbitrary affine equivariant functions. Then, whenever the expectations (which are with respect to ε) exist,*

$$\begin{aligned} E(\hat{\mu}(y)) &= \eta \\ E(\hat{\Sigma}(y)) &= c_0 \Omega \\ \mathrm{Cov}(\hat{\mu}(y), \hat{\mu}(y)) &= c_1 \Omega \\ \mathrm{Cov}(\hat{\mu}(y), \hat{\Sigma}(y)) &= 0 \\ \mathrm{Cov}(\hat{\Sigma}(y), \hat{\Sigma}(y)) &= c_2 \Omega^{\otimes 2} + c_3(\Omega^{\otimes 2*} + \Omega^{\otimes 2**}), \end{aligned} \tag{4.1}$$

where $\Omega^{\otimes 2}$, $\Omega^{\otimes 2*}$, and $\Omega^{\otimes 2**}$ are described in Theorem 3.1.

As mentioned above, this result has been derived many times. The extension to third and fourth cumulants given in the next theorem is new. It demonstrates a virtue of the tensor approach to this problem, which is that extensions are no more difficult than the second-order analysis.

COROLLARY 4.2. *Under the assumptions of Corollary 4.1 the third cumulants of $\hat{\mu}$ and $\hat{\Sigma}$ are given by*

$$\begin{aligned} \kappa_{\mu\mu\mu} &= 0 \\ \kappa_{\mu\mu\Sigma} &= c_4 \Omega^{\otimes 2} + c_5(\Omega^{\otimes 2*} + \Omega^{\otimes 2**}) \\ \kappa_{\mu\Sigma\Sigma} &= 0 \end{aligned} \tag{4.2}$$

$$\kappa_{\Sigma\Sigma\Sigma} = c_6 \Omega^{\otimes 3} + c_7 \sum_{\omega \vee \omega_0 < 1} (\tau^{\omega})^{\Omega^{1/2}} + c_8 \sum_{\omega \vee \omega_0 = 1} (\tau^{\omega})^{\Omega^{1/2}},$$

where $\omega_0 = (12)(34)(56)$ *and* $\Omega^{\otimes 3}$ *is defined in Theorem* 3.3; *the* c_i*'s are constants. The fourth cumulant tensors are given by*

$$\kappa_{\mu\mu\mu\mu} = c_9 \Omega^{\otimes 2} + c_{10}(\Omega^{\otimes 2*} + \Omega^{\otimes 2**})$$

$$\kappa_{\mu\mu\mu\Sigma} = 0$$

$$\kappa_{\mu\mu\Sigma\Sigma} = c_{11}\Omega^{\otimes 3} + c_{12} \sum_{\omega \vee \omega_0 = \omega_1} (\tau^{\omega})^{\Omega^{1/2}} + c_{13} \sum_{\omega \vee \omega_0 \in \{\omega_2, \omega_3\}} (\tau^{\omega})^{\Omega^{1/2}} + c_{14} \sum_{\omega \vee \omega_0 = 1} (\tau^{\omega})^{\Omega^{1/2}}$$

$$\kappa_{\mu\Sigma\Sigma\Sigma} = 0$$

$$\kappa_{\Sigma\Sigma\Sigma\Sigma} = c_{15}\Omega^{\otimes 4} + c_{16} \sum_{\nu(\omega \vee \omega_0) = 3} (\tau^{\omega})^{\Omega^{1/2}} + c_{17} \sum_{\nu(\omega \vee \omega_0) = 2} (\tau^{\omega})^{\Omega^{1/2}} + c_{18} \sum_{\omega \vee \omega_0 = 1} (\tau^{\omega})^{\Omega^{1/2}}, \tag{4.3}$$

where $\omega_0 = (12)(34)(56)(78)$, $\omega_1 = (12)(3456)$, $\omega_2 = (1234)(56)$, $\omega_3 = (1256)(34)$, *and* $\Omega^{\otimes 4}$ *is defined in Theorem* 3.3; *the* c_i*'s are constants.*

The methods in this paper can also be used to study equivariant functions other than mean and scatter matrix estimators. One important example of this involves estimates of multivariate third and fourth moments or cumulants.

THEOREM 4.1. *Under the assumptions of Corollary* 4.1, *let* $\hat{K}_3 : \mathbf{R}^d \otimes \mathbf{R}^n \mapsto \mathbf{R}^d \otimes \mathbf{R}^d \otimes \mathbf{R}^d$ *and* $\hat{K}_4 : \mathbf{R}^d \otimes \mathbf{R}^n \mapsto \mathbf{R}^d \otimes \mathbf{R}^d \otimes \mathbf{R}^d \otimes \mathbf{R}^d$ *be arbitrary equivariant functions. Then*

$$E(K_3(y)) = 0$$

$$E(\hat{K}_4(y)) = c_{19}\Omega^{\otimes 2} + c_{20}(\Omega^{\otimes 2*} + \Omega^{\otimes 2**})$$

$$\operatorname{Cov}(\hat{K}_3(y), \hat{K}_3(y)) = c_{21}\Omega^{\otimes 3} + c_{22} \sum_{\omega \wedge \omega' \neq 0} (\tau^{\omega})^{\Omega^{1/2}} + c_{23} \sum_{\omega \wedge \omega' = 0} (\tau^{\omega})^{\Omega^{1/2}}$$

$$\operatorname{Cov}(\hat{K}_3(y), \hat{K}_4(y)) = 0$$

$$\operatorname{Cov}(\hat{K}_4(y), \hat{K}_4(y)) = c_{24}\Omega^{\otimes 4} + c_{25} \sum_{\nu(\omega \wedge \omega'') = 4} (\tau^{\omega})^{\Omega^{1/2}} + c_{26} \sum_{\nu(\omega \wedge \omega'') = 6} (\tau^{\omega})^{\Omega^{1/2}} + c_{27} \sum_{\omega \wedge \omega'' = 0} (\tau^{\omega})^{\Omega^{1/2}} \tag{4.4}$$

where $\omega' = (123)(456)$, *and* $\omega'' = (1234)(5678)$; *the* c_i*'s are constants again, and where* $\omega \wedge \omega''$ *is the lattice meet.*

One class of examples for this problem is given by the M-estimates of multivariate location and scatter. These estimators are given by the solution of two sets of estimating equations (see [9, 10, 14, 26–28]). The estimating equations for location are similar to maximum likelihood for an elliptical family (see Section 4.2 below), but the estimating equations for scatter are usually chosen to provide greater robustness against outliers than would be obtained by using the ones from the elliptical case. Other estimators that fit the assumptions include the multivariate generalizations of the median and trimmed mean [6, 7], the minimum volume ellipsoid estimator [21, 23], related S-estimators [4], and the minimum covariance determinant estimator [22].

4.2. *Maximum Likelihood for Elliptical Families*

Suppose each d-dimensional observation is elliptically distributed; that is $y_i \sim Ell_{d,h}(\eta, \Omega)$. Then the log likelihood for the entire sample y is given by

$$L(\eta, \Omega; y) = \frac{-n}{2} \ln(|\Omega|) + \sum_i h((y_i - \eta)^{\mathrm{T}} \Omega^{-1}(y_i - \eta)) \tag{4.7}$$

and the maximum likelihood estimators (MLE) $\hat{\eta}$ and $\hat{\Omega}$ satisfy

$$0 = L_\eta = 2 \sum_i g_i \Omega^{-1}(y_i - \hat{\eta}) \tag{4.8}$$

$$0 = L_\Omega = -\frac{2}{n} \hat{\Omega}^{-1} - \sum_i g_i \Omega^{-1}(y_i - \hat{\eta})(y_i - \hat{\eta})^{\mathrm{T}} \hat{\Omega}^{-1}, \tag{4.9}$$

where $g_i = g((y_i - \hat{\eta})^{\mathrm{T}} \hat{\Omega}^{-1}(y_i - \hat{\eta}))$ with $g(x) = d \ln(h(x))/dx = h'(x)/h(x)$. We denote the full parameter by $\theta = (\eta, \Omega) \in \mathbf{R}^d \times \mathrm{PSD}(d)$. It is easy to check that the expected information tensor $E(L_\theta \otimes L_\theta) = E(L_\eta \otimes L_\eta) \times E(L_\eta \otimes L_\Omega) \times E(L_\Omega \otimes L_\eta) \times E(L_\Omega \otimes L_\Omega)$ consists of four orthogonal invariant tensors of degrees 2, 3, 3, and 4, respectively. Thus the structure of the information matrix is given by Theorem 3.1. The coefficients in this case have been given by Mitchell [17] in terms of the function $h(\cdot)$. After inversion, this gives the asymptotic values of the coefficients in Theorem 3.1—the finite sample values would need to be determined by simulation. Even in the normal case, the fact that $\hat{\Omega}$ us is not unbiased for Ω means that the coefficients would differ by terms of order $1/\eta$. However, if the MLE of Ω were redefined by multiplying by $n/(n-1)$, then exact equality holds here.

4.3. *Multivariate Regression*

Suppose for each $1 \leqslant i \leqslant n$, one has a set x_{ri}, $1 \leqslant r \leqslant p$, of carriers that are either fixed or random independent of ε. If the carriers are random, then the covariances given below are understood to be conditional on the achieved values of the carriers. Let $x = \sum x_{ri} e_r \otimes e_i \in \mathbf{R}^p \otimes \mathbf{R}^n$. Suppose that the model is $y = \beta x + \varepsilon$, where $\beta \in \mathbf{R}^d \otimes \mathbf{R}^p$. Then the subspace $V = \sum e_s \otimes \tilde{V} \subset \mathbf{R}^n$ is generated by the p n-vectors x_r. Assuming that $\varepsilon_i \sim$ i.i.d. $Ell_{d,h}(0, \Omega)$ and that one has equivariant estimators $\hat{\mu}$ and $\hat{\Sigma}$, then $\hat{\mu}$ estimates βx, and (possibly nonunique) values for $\hat{\beta}$ can be derived by solving the linear equations involved. Thus $\hat{\beta} = \hat{\mu} x^{\mathrm{T}} (x x^{\mathrm{T}})^{-1}$ (a generalized inverse is used if $x x^{\mathrm{T}}$ is singular). The covariance matrix of $\hat{\beta}$ satisfies $\mathrm{Cov}(\hat{\mu}) = \mathrm{Cov}(\hat{\beta} x) = x^{\mathrm{T}} \, \mathrm{Cov}(\hat{\beta}) \, x$, so that

$$\mathrm{Cov}(\hat{\beta} x) = \Omega \otimes (x x^{\mathrm{T}})^{-1} x C_1 x^{\mathrm{T}} (x x^{\mathrm{T}})^{-1}. \tag{4.10}$$

If the estimator is defined so that $c_0 = 1$, then clearly the substitution of $\hat{\Sigma}$ for Ω in this expression provides an unbiased estimate of the covariance matrix (assuming that C_1 is known).

In regression, the undetermined constants in the covariance matrix of $\hat{\mu}$ have been reduced from an $n \times n$ matrix to a $p \times p$ matrix. Although this still may seem like a great number of constants, it must be understood that the class of estimators to which the result applies is very large. For example, if one has n different equivariant regression procedures, then one may use a different one to produce the prediction for each data point. It should not be expected that, in such a case, the covariance matrix should always be a multiple of $(x x^{\mathrm{T}})^{-1}$. In any case, Theorem 3.2 shows that there exists a linear equivariant estimator to match each choice of $\mathrm{Cov}(\hat{\beta})$.

This result is known for many classes of estimators. The advantage of the present formulation is both the generality and the ability to extend the results to a higher order. If one were concerned, for example, with the convergence of $\hat{\Sigma}$ to multivariate normality, the third-order cumulant tensor would be relevant—this is given up to three constants in Corollary 4.2, and these may be determined by simulation or approximated by asymptotic analysis.

This analysis can also be expanded to cover the case of seemingly unrelated regression [30]. The setup here is identical to that of multivariate regression, except that the carriers used to predict each of the d components of y_i are possibly different, so that the subspaces of $\mathbf{R}^n$ into which the d components of y are projected are different. The maximum likelihood estimator is clearly affine equivariant but no estimator for this problem satisfies the conditions of Theorem 3.2. Hence, there is no linear approximation to any such estimator that does not depend on y. In the usual formulation, let Y be an $nd \times 1$ vector which consists of the d response vectors stacked vertically (and similarly for the d residual n-vectors r_s and the dp_s-vectors β_s). Let X be a block diagonal matrix whose sth block is X_s, the matrix of carriers for the sth regression. Then the MLE satisfies

$$\begin{aligned} \hat{\beta} &= (X^{\mathrm{T}}\hat{T}X)^{-1} X^{\mathrm{T}}\hat{T}Y \\ \hat{T} &= (\hat{\Sigma} \otimes I)^{-1} \\ \hat{\sigma}_{st} &= n^{-1} r_s^{\mathrm{T}} r_t. \end{aligned} \tag{4.11}$$

This estimator can be shown to be affine equivariant, but the estimator is only linear conditional on T, whose value depends on y. It is also of interest that Zellner's [30] one-step approximation is not affine equivariant, since it depends on the coordinatization used to start the iterations.

5. Conclusions

In this paper, we have derived results that considerably restrict possible values for the cumulant tensors of equivariant estimators for statistical models with elliptical errors. In many cases there are only a few undetermined constants, the remaining structure being entirely determined by affine equivariance and elliptical errors. The method is particularly effective for considering equivariant estimators of scatter as well as third and fourth cumulant tensors, since the number of undetermined constants is often quite small. For example, the covariance tensor of $\hat{\Sigma}$ is determined up to two unknown constants. Multivariate location and multivariate regression are other applications for which the results should be useful. In future work, we intend to extend the results to other models; for example, multivariate time series models may fail to satisfy the assumptions, since the expected value of an observation will usually depend on the errors from previous observations.

References

[1] Anderson, T. W., Fang, K. T., and Hsu, H. (1986). Maximum likelihood estimates and likelihood ratio criteria for multivariate elliptically contoured distributions. *Canad. J. Statist.* **14** 55–59.

[2] Chmielewski, M. A. (1980). Invariant scale matrix hypothesis tests under elliptical symmetry. *J. Multivariate Anal.* **10** 343–350.

[3] Chmielewski, M. A. (1981). Elliptically symmetric distributions: A review and bibliography. *Internat. Statist. Rev.* **49** 67–74.

[4] Davies, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

[5] Devlin, S. J., Gnanadesikan, R., and Kettenring, J. R. (1976). Some multivariate applications of elliptical distributions. In *Essays in Probability and Statistics* (S. Ikeda, Ed.). Shinko Tsusho, Tokyo.

[6] DONOHO, D. L. (1982). *Breakdown Properties of Multivariate Location Estimators*, Ph. D. Qualifying Paper, Department of Statistics, Harvard University.

[7] DONOHO, D. L., AND GASKO, M. (1987). *Multivariate Generalizations of the Median and Trimmed Mean, I.* Technical Report 133, Department of Statistics, University of California, Berkeley.

[8] EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* Wiley, New York.

[9] HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J., AND STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[10] HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

[11] JEFFREYS, H. (1952). *Cartesian Tensors.* Cambridge Univ. Press, Cambridge.

[12] KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization, *Sankhyā Ser. A* **32** 419–430.

[13] KENDALL, M., AND STUART, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. Macmillan Co., New York.

[14] MARONNA, R. A. (1976). Robust *M*-estimators of multivariate location and scatter, *Ann. Statist.* **4** 51–67.

[15] MCCULLAGH, P. (1987). *Tensor Methods in Statistics.* Chapman & Hall, London.

[16] MALLOWS, C. L. (1961). Latent vectors of random symmetric matrices. *Biometrika* **48** 133–149.

[17] MITCHELL, A. F. S. (1989). The information matrix, skewness tensor and α-connections for the general multivariate elliptic distribution. *Ann. Inst. Statist. Math.* **41** 289–304.

[18] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

[19] MUIRHEAD, R. J., AND WATERNAUX, C. M. (1980). Asymptotic distributions in canonical corqelation analysis and other multivariate procedures for nonnormal populations. *Biometrika* **67** 31–43.

[20] PIERCE, R. S. (1982). *Associative Algebras.* Springer, New York.

[21] ROUSSEEUW, P. J. (1984). Least median of squares regression, *J. Amer. Statist. Assoc.* **79** 871–880.

[22] ROUSSEEUW, P. J. (1986). Multivariate estimation with high breakdown point. In *Proceedings, 4th Pannonian Symposium on Mathematical Statistics, Bad Tatzmannsdorf* (W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds.), pp. 283–297. Reidel, Dordrecht.

[23] ROUSSEEUW, P. J., AND LEROY, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

[24] SHAPIRO, A., AND BROWNE, M. (1987). Analysis of covariance structures under elliptical distributions, *J. Amer. Statist. Assoc.* **82** 1092–1097.

[25] THOMAS, T. Y. (1965). *Concepts from Tensor Analysis and Differential Geometry.* Academic Press, New York.

[26] TYLER, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika* **69** 429–436.

[27] TYLER, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika* **70** 411–420.

[28] TYLER, D. E. (1987). A distribution-free *M*-estimator of multivariate scatter, *Ann. Statist.* **15** 234–251.

[29] YOHAI, V. J., AND MARONNA, R. A. (1979). Asymptotic behavior of *M*-estimators for the linear model. *Ann. Statist.* **7** 258–268.

[30] ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *J. Amer. Statist. Assoc.* **57** 348–368.