Metrika, Volume 27, 1980, page 103–113. ©Physica-Verlag, Vienna.

Multilinear Estimation of Skewness and Kurtosis in Linear Models

By *F. Pukelsheim*, Freiburg im Breisgau and Stanford[1])

Summary: Estimation of the coefficients of skewness and kurtosis in a classical linear model situation is presented as an application of multilinear algebra and standard theory of mean estimation. The resulting estimators have optimality properties among all estimators that are invariant under mean translations, polynomials of degree three (skewness) or four (kurtosis) in the observations, and unbiased.

1. Introduction

Consider a *classical linear model*

$$Y \sim (X\beta; \sigma^2 I_n) \tag{1.1}$$

in whose definition we include the assumptions that the components of the random $\mathbf{R}^n$-vector Y are independent, and have common coefficients γ_1 of skewness and γ_2 of kurtosis. In other words, Y has independent components, and $\mathcal{E}Y = X\beta$, $\mathcal{D}Y = \sigma^2 I_n$, and for all $i = 1, \ldots, n$,

$$\mathcal{E}\left[\frac{Y_i - \mathcal{E}Y_i}{\sqrt{\operatorname{Var} Y_i}}\right]^3 = \gamma_1; \; \mathcal{E}\left[\frac{Y_i - \mathcal{E}Y_i}{\sqrt{\operatorname{Var} Y_i}}\right]^4 - 3 = \gamma_2. \tag{1.2}$$

A distinguished case satisfying these requirements, and hence possibly underlying model (1.1), is the normal law, i.e., $Y \sim N_n (X\beta; \sigma^2 I_n)$.

The purpose of this paper is to present polynomial estimators for skewness (Section 3) and kurtosis (Section 4) which are unbiased under every distribution that fulfills the assumptions of model (1.1), and which are of minimum variance under normality. The concept of unbiased estimation with minimum variance at a distinguished parameter value is of particular interest in linear model theory. For a more general context see the textbook of *Schmetterer* [1974, p. 273]. It seems to me that there are three major grounds to justify this interest: (i) In a theoretical exposition, investigation of unbiased estimators with local minimum variance properties provides a possible approach to *uniform* minimum variance unbiased estimation procedures. (ii) In cases

---

[1]) *Friedrich Pukelsheim*, Institut für Mathematische Stochastik der Universität, Herman Herder Str. 10, D–7800 Freiburg im Breisgau.

when uniformly best procedures do not exist unbiased estimation with local minimum variance properties may be an applicable procedure, possibly iterative. (iii) For testing hypotheses, special interest is directed toward the distribution of the test statistic under the nullhypothesis, i.e., under one (or more) distinguished parameter values: local minimum variance properties of an estimator provide information in this direction.

When estimating second and higher moments, a natural requirement is that of location invariance, or equivalently, that the estimates depend on the observations Y only through the *residual statistic*

$$Z := MY, \; M := I_n - XX^{+}, \tag{1.3}$$

where M projects orthogonally onto the orthogonal complement of the range of X. Z is the vector of residuals obtained from a simple least squares fit for β, and this statistic is also maximal invariant with respect to the translation group $\{y \to y + Xb \mid b \in \mathbf{R}^p\}$, where the matrix X is assumed to be of order $n \times p$. Estimating the third moment coefficient γ_1, we shall restrict attention to the class

- Γ_1 of all homogeneous polynomials of degree three in the residuals $Z_1, \ldots, Z_n$, and to its subclass
- G_1 with pure cubes $Z_1^3, \ldots, Z_n^3$ only. As for the fourth moment coefficient γ_2, we analogously investigate the class
- Γ_2 of all constants plus homogeneous polynomials of degree four in $Z_1, \ldots, Z_n$, and its subclass
- G_2 of constants plus polynomials with pure fourth powers $Z_1^4, \ldots, Z_n^4$ only.

In Section 2 the Kronecker and Hadamard third and fourth powers of the residual vector Z yield *derived models* in which (i) the unknown coefficient γ_i appears as a mean parameter, and (ii) Γ_i and G_i turn out to be the classes of all *linear* estimators; $i = 1, 2$. Standard theory of mean estimation, then, is all that is needed for finding unbiased estimators of skewness (Section 3) and kurtosis (Section 4) which, under normality, are of minimum variance. The technical difficulties reduce to appropriately displaying the moments of the residual statistic Z. We shall use two tools to facilitate this task, the minor one of which is the use of the vec operator: this allows to avoid summation signs through expressions like $\sum_{i=1}^{n} e_i \otimes e_i = \operatorname{vec} I_n$. A major tool is the second one, the *symmetrizer* π_s. For example, if U is $N_n\,(0; I_n)$ – distributed and the sixth moments are to be computed, then $U_{\sigma(1)}^4\, U_{\sigma(2)}^2$ has expectation 3 regardless of the particular permutation σ of the components $U_1, \ldots, U_n$; and this suggests a two-stage approach: In the first step expectations of U_1^6, $U_1^4\, U_2^2$, $U_1^2\, U_2^2\, U_3^2$ are computed, and in the second step the symmetrizer π_s is employed in order to generate the necessary permutations. It is found that these means greatly simplify the representation of the covariance operators in the derived models.

Our approach is a natural extension of the *dispersion mean correspondence* that reduces the estimation of variance covariance components to standard mean estimation [*Pukelsheim*]. When estimating skewness and kurtosis, it would appear natural to also insist on scale invariance. This, however, excludes the use of estimators which are mere polynomials of the observations, and hence surpasses means and methods of multilinear algebra. Hence we sacrifice scale invariance to theoretical elegance, and neglected scale invariance does, indeed, secure a manageable theory. But it also results in estimators that depend on the unknown value of σ^2. There are, of course, possibilities of various degrees of arbitrariness of how to replace σ^2 by an estimate $\hat{\sigma}^2$, and some courses of action will be discussed in the examples. This discussion will show that while the proposed multilinear approach yields procedures of direct applicability only for the genuine third and fourth central moments, it nevertheless motivates a guideline of how estimators for the coefficients of skewness and kurtosis may be obtained as well.

The examples are taken from *Anscombe's* paper [1961] who used estimates of skewness and kurtosis to study departure from normality.

2. The Derived Models

Extensive use will be made of the Kronecker product $A \otimes B := [[A_{ij}B]]$ for two matrices A and B of arbitrary order, and of the Hadamard product $A * B := ((A_{ij}B_{ij}))$ for matrices A and B of the same order, see *Rao* [1973, pp. 29–30], *Styan* [1973]. Powers are written as $A^{\otimes 2} := A \otimes A$, $A^{*2} := A * A$, etc.; recall the fundamental rule $(A \otimes B)(C \otimes D) = AC \otimes BD$.

The components of the third Kronecker power $Z^{\otimes 3}$ furnish the n^3 mixed third powers of $Z_1, \ldots, Z_n$, whence any homogeneous polynomial of degree three may be written as a linear form $w'Z^{\otimes 3}$. The fewer number of components of the third Hadamard power Z^{*3} suffice for investigation whenever the polynomial involves pure cubes $Z_1^3, \ldots, Z_n^3$, only:

$$\Gamma_1 = \{w'Z^{\otimes 3} \mid w \in \mathbf{R}^{n^3}\}, \quad G_1 = \{w'Z^{*3} \mid w \in \mathbf{R}^n\}. \tag{2.1}$$

Hence we study the *derived models* generated by $Z^{\otimes 3}$ and Z^{*3}. Computations are facilitated by the fact that $Z^{*3} = H_3 Z^{\otimes 3}$ is a linear transformation of $Z^{\otimes 3}$, where

$$H_3 := \sum_{i=1}^{n} e_i (e_i^{\otimes 3})' \tag{2.2}$$

is a $n \times n^3$ matrix; e_i is the i-th Euclidean basis vector in $\mathbf{R}^n$, with i-th component equal to one and zeroes elsewhere. Analogously, define the $n \times n^4$ matrix $H_4 = \sum_{i=1}^{n} e_i (e_i^{\otimes 4})'$. We shall also use $1_n := \Sigma e_i = (1, \ldots, 1)'$. The following lemma shows that both $Z^{\otimes 3}$ and Z^{*3} give rise to models whose expection is linear in $\sigma^3 \gamma_1$.

Lemma 2.1: For every classical linear model:

a) $EZ^{\otimes 3} = \sum_{i=1}^{n} (Me_i)^{\otimes 3} \cdot \sigma^3 \gamma_1$.

b) $EZ^{*3} = M^{*3} 1_n \, \sigma^3 \gamma_1$.

c) There exists an unbiased estimator for γ_1 in Γ_1 or in G_1 if and only if $1_n' M^{*3} 1_n > 0$.

Proof: The independent components of $U := \sigma^{-1}(Y - EY)$ have mean zero, variance one, and coefficient of skewness γ_1. Hence the expected value vanishes for powers like $U_1 U_2 U_3$ and $U_1 U_2^2$, and equals γ_1 for U_1^3. a) then follows from $EZ^{\otimes 3} = \sigma^3 M^{\otimes 3}\ EU^{\otimes 3}$; and applying H_3 in a) yields b). As for c), Γ_1-estimability is ensured whenever $\Sigma\,(Me_i)^{\otimes 3} \neq 0$, i.e., $\|\Sigma\,(Me_i)^{\otimes 3}\|^2 = \Sigma\,(e_i' M^2 e_j)^3 = \Sigma\, e_i' M^{*3} e_j > 0$. G_1-estimability requires $M^{*3}\,1_n \neq 0$. The assertion then follows from the nonnegative-definiteness of M^{*3} [*Styan*, p. 221]. □

Before proceeding to higher powers we introduce the *symmetrizer* π_s [*Greub*, pp.90f.]. As an endomorphism of a p-fold Kronecker power $\otimes^p \mathbf{R}^n$, it suffices to have π_s defined on the spanning set of all p-fold Kronecker powers $x_1 \otimes \ldots \otimes x_p$ of p $\mathbf{R}^n$-vectors $x_1, \ldots, x_p$ by

$$\pi_s\,(x_1 \otimes \ldots \otimes x_p) := \frac{1}{p!}\,\Sigma_{\sigma \in S_p}\,(x_{\sigma(1)} \otimes \ldots \otimes x_{\sigma(p)}) \tag{2.3}$$

where S_p is the symmetric group (permutation group) of order p. A point in $\otimes^p \mathbf{R}^n$ is called *symmetric* if it lies in the range of π_s. A mapping L defined on $\otimes^p \mathbf{R}^n$ is called *symmetric* if $L = L \circ \pi_s$.

For example, consider the case $p = 2$. Then $\pi_s\,(x \otimes y) = \frac{1}{2}\,(x \otimes y + y \otimes x)$. Using the isomorphism vec: $\mathbf{R}^{n \times n} \to \mathbf{R}^{n^2}$, $xy' \to x \otimes y$, the result of π_s on the space $\mathbf{R}^{n \times n}$ of $n \times n$ matrices is exactly that it projects a square matrix A into its symmetric part $\frac{1}{2}(A + A')$; further properties of the vec operator are $\sum_{i=1}^{n} e_i \otimes e_i = \mathrm{vec}\, I_n$, and $\mathrm{vec}\, ABC = A \otimes C' \cdot \mathrm{vec}\, B$ [*Pukelsheim*, p. 326]. – Notice that $z^{\otimes 3}$ is symmetric; hence in a linear form $w'z^{\otimes 3}$ we may assume w to be symmetric, too. Symmetric mappings are those given by H_3, or $\Sigma\,(e_i' M)^{\otimes 3}$.

The classes Γ_2 and G_2 of estimators are now represented by using the fourth Kronecker and Hadamard powers $Z^{\otimes 4}$ and Z^{*4}:

$$\Gamma_2 = \{w'Z^{\otimes 4} + c_\Gamma \mid w \in \mathbf{R}^{n^4},\ c_\Gamma \in \mathbf{R}\},$$

$$G_2 = \{w'Z^{*4} + c_G \mid w \in \mathbf{R}^n,\ c_G \in \mathbf{R}\}.$$

Analogously to Lemma 2.1 we next investigate the expectations of $Z^{\otimes 4}$ and Z^{*4}. To this end, define

$$m := (M_{11}, \ldots, M_{nn})'. \tag{2.4}$$

Lemma 2.2: For every classical linear model:

a) $EZ^{\otimes 4} = \sum_{i=1}^{n} (Me_i)^{\otimes 4} \cdot \sigma^4 \gamma_2 + 3\sigma^4\, \pi_s\,(\mathrm{vec}\, M)^{\otimes 2}$.

b) $EZ^{*4} = M^{*4}\,1_n\, \sigma^4 \gamma_2 + 3\sigma^4 m^{*2}$.

c) There exists an unbiased estimator for γ_2 in Γ_2 or in G_2 if and only if the rank of X is smaller than n.

Proof: Again use the standardized U as in the proof of Lemma 2.1 The expected values of $U_1 U_2 U_3 U_4$, $U_1 U_2 U_3^2$, $U_1 U_2^3$ vanish, leaving the cases $U_1^2 U_2^2$, and U_1^4. From (1.2), the latter has expectation $EU_1^4 = \gamma_2 + 3$. The former has expectation 1 and may arise from $U_1 U_1 U_2 U_2$, $U_1 U_2 U_1 U_2$, or $U_1 U_2 U_2 U_1$. This gives $EU^{\otimes 4} = \Sigma e_i^{\otimes 4} \cdot (\gamma_2 + 3) + \Sigma_{i \neq j} (e_i \otimes e_i \otimes e_j \otimes e_j + e_i \otimes e_j \otimes e_i \otimes e_j + e_i \otimes e_j \otimes e_j \otimes e_i) = \Sigma e_i^{\otimes 4} \cdot \gamma_2 + \Sigma_{i,j} (e_i \otimes e_i \otimes e_j \otimes e_j + e_i \otimes e_j \otimes e_i \otimes e_j + e_i \otimes e_j \otimes e_j \otimes e_i)$. The last term equals $3\pi_s \cdot \Sigma_{i,j} e_i \otimes e_i \otimes e_j \otimes e_j$, and this proves a). b) results when applying the transformation H_4 in a). In c) we first get $\Sigma M_{ij}^4 \neq 0$; but this is the same as $M \neq 0$, i.e., range $X \neq \mathbf{R}^n$. □

A classical linear model tacitly assumes that the rank of X is less than n, hence γ_2 is "always" unbiasedly estimable. Lemma 2.2 also provides the values for the constants c_Γ and c_G:

Corollary 2.3: For the classes Γ_2 and G_2, unbiasedness of $w'Z^{\otimes 4} + c_\Gamma$ $(w \in \mathbf{R}^{n^4}, c_\Gamma \in \mathbf{R})$ for γ_2 implies

$$c_\Gamma = -3\sigma^4 w'\pi_s \cdot (\operatorname{vec} M)^{\otimes 2},$$

and unbiasedness of $w'Z^{*4} + c_G$ $(w \in \mathbf{R}^n, c_G \in \mathbf{R})$ for γ_2 implies

$$c_G = -3\sigma^4 w'm^{*2}. \quad \square$$

This section is concluded with some upper and lower bounds associated with the projector M. Let m be as in (2.4); define p to be rank of X, and ν to be the rank of M, so that $p + \nu = n$.

Lemma 2.4:

a) For every $k = 1, 2, \ldots$, one has $0 \leqslant m'M^{*k}m \leqslant \nu$ and $\nu^k / n^{k-1} \leqslant \sum_{i=1}^{n} M_{ii}^k$.

b) One has $\sum_{i,j=1}^{n} M_{ij}^3 \geqslant \nu(\nu^2 - np)/n^2$, and $\nu^2 - np$ is positive if and only if $n > (3 + \sqrt{5})p/2$.

Proof: a) Since M is a projector, it follows from Theorem 3.1 and Corollary 3.2 in *Styan* [1973] that the eigenvalues of M^{*k} lie in the unit interval [0, 1], whence $0 \leqslant m'M^{*k}m \leqslant m'm$. But

$$M_{ii} = \sum_{j=1}^{n} M_{ij}^2 \tag{2.5}$$

implies $M_{ii} \in [0, 1]$ as well, so that $m'm \leqslant \Sigma M_{ii} = \nu$. The second part is Jensen's Inequality: $\nu^k / n^k = (\Sigma M_{ii}/n)^k \leqslant \Sigma M_{ii}^k / n$.

b) Certainly, $\Sigma M_{ij}^3 \geqslant \Sigma M_{ii}^3 - \Sigma_{i \neq j} |M_{ij}| M_{ij}^2$. But (2.5) implies $M_{ij}^2 \leqslant |M_{ij}|$, and $-\Sigma_{i \neq j} |M_{ij}|^2 = M_{jj}^2 - M_{ii}$. This gives $\Sigma M_{ij}^3 \geqslant \Sigma M_{ii}^3 + \Sigma M_{ii}^2 - \Sigma M_{ii}$. Now use a) to obtain the first assertion.

For fixed p, the zeros of $x^2 - 3xp + p^2$ are $(3 \pm \sqrt{5})\, p \,/\, 2$. Noting that $n \geqslant p$ finishes the proof. □

The lower bound $\nu\,(\nu^2 - np) \,/\, n^2$ for $\Sigma\, M_{ij}^3$ is rather crude since it is not always nonnegative. Yet for the estimability criterion in Lemma 2.1 c), one obtains the sufficient condition $n \geqslant 3p$. The case of two observations with a homogeneous mean provides an example for $n = 2p$ and $\Sigma\, M_{ij}^3 = 0$.

Next we compute the sixth and eighth moments under normality, and then turn to minimum variance estimation.

3. Skewness

In the sequel we speak of a *quasinormal* model whenever the assumption of normality is used only to compute the moments involved.

In order to display higher moments we shall use the (i, j)-th Euclidean basis matrix $E_{ij} := e_i e_j'$ that carries a one on place (i, j) and zeros elsewhere, and the diagonal matrix $\mathrm{Diag}\, M$ whose diagonal is copied from M.

Lemma 3.1: For every quasinormal classical linear model the dispersion matrices of $Z^{\otimes 3}$ and Z^{*3} are given by $\sigma^6 \pi_s\, S_\Gamma\, \pi_s$ and $\sigma^6\, S_G$, respectively, where

$$S_\Gamma := 6\, M^{\otimes 3} + 9\, M \otimes (\,\mathrm{vec}\, M \cdot \mathrm{vec}' M)$$

$$S_G := 6\, M^{*3} + 9\,(\mathrm{Diag}\, M)\, M\, (\mathrm{Diag}\, M).$$

Proof: Define $S := 6\, I_n^{\otimes 3} + 9 \sum_{i,j,k=1}^{n} E_{ii} \otimes E_{jk} \otimes E_{jk}$. It suffices to show that for a $N_n\,(0; I_n)$-distributed U one has $\mathcal{D} U^{\otimes 3} = \pi_s S \pi_s$. Since $M^{\otimes 3}$ and π_s commute, this implies $\mathcal{D} Z^{\otimes 3} = \sigma^6\, M^{\otimes 3}\, (\mathcal{D} U^{\otimes 3})\, M^{\otimes 3} = \sigma^6 \pi_s \cdot M^{\otimes 3} S M^{\otimes 3} \cdot \pi_s$. Using $\mathrm{vec}\, ABC = A \otimes C' \cdot \mathrm{vec}\, B$ one obtains $M \otimes M^{\otimes 2}\, [6\, I_n^{\otimes 3} + 9\, I_n \otimes (\mathrm{vec}\, I_n \cdot \mathrm{vec}'\, I_n)]\, M \otimes M^{\otimes 2} = S_\Gamma$. Employing H_3, one also gets $\mathcal{D} Z^{*3} = \sigma^6\, H_3\, S_\Gamma\, H_3' = \sigma^6\, S_G$, as asserted.

Note that $\mathcal{E} U^{\otimes 3} = 0$. The equality $\mathcal{E}\,(U^{\otimes 3})\,(U^{\otimes 3})' = \pi_s S \pi_s$ is now verified elementwise. Let $\alpha, \beta, \gamma = 1, \ldots, n$ be pairwise different. The only nonvanishing sixth moments arise from U_α^6 with expectation 15, and $U_\alpha^4 U_\beta^2$ with expectation 3, and $U_\alpha^2 U_\beta^2\, U_\gamma^2$ with expectation 1. In the first case, $(e_\alpha^{\otimes 3})'\, \pi_s\, S \pi_s\, (e_\alpha^{\otimes 3}) = 15$. For the second case use $\pi_s\,(e_\alpha \otimes e_\beta \otimes e_\beta) = (e_\alpha \otimes e_\beta \otimes e_\beta + e_\beta \otimes e_\alpha \otimes e_\beta + e_\beta \otimes e_\beta \otimes e_\alpha) \,/\, 3$, and obtain $(e_\alpha^{\otimes 3})'\, \pi_s S \pi_s\, (e_\alpha \otimes e_\beta \otimes e_\beta) = 3$ and $(e_\alpha \otimes e_\alpha \otimes e_\beta)'\, \pi_s\, S \pi_s\, (e_\alpha \otimes e_\alpha \otimes e_\beta) = 3$. Along the same lines it is found that $(e_\alpha \otimes e_\alpha \otimes e_\beta)' \pi_s S \pi_s\, (e_\beta \otimes e_\gamma \otimes e_\gamma) = 1$, and, finally, $(e_\alpha \otimes e_\beta \otimes e_\gamma)' \pi_s S \pi_s\, (e_\alpha \otimes e_\beta \otimes e_\gamma) = 1$. For vanishing moments of order six the corresponding entry of $\pi_s S \pi_s$ vanishes, too. □

Note that S_G is also given by equation (15) in *Anscombe* [1961].

In case σ^2 is known, we are now in a position to exhibit various estimators of the coefficient of skewness γ_1, and discuss their relationship.

Theorem 1: Assume a classical linear model $Y \sim (X\beta; \sigma^2 I_n)$ such that $1_n' M^{*3} 1_n > 0$.

a) The Aitken estimate for γ_1 in the derived model

$$Z^{\otimes 3} \sim (\sum_{i=1}^{n} (Me_i)^{\otimes 3} \cdot \sigma^3 \gamma_1 ; \sigma^6 \pi_s S_\Gamma \pi_s)$$

is given by

$$\hat{\gamma}_1 := (\sigma^3 \sum_{i,j=1}^{n} e_i'^{\otimes 3} (\pi_s S_\Gamma \pi_s)^+ e_j^{\otimes 3})^{-1} \sum_{i=1}^{n} e_i'^{\otimes 3} (\pi_s S_\Gamma \pi_s)^+ Z^{\otimes 3}.$$

$\hat{\gamma}_1$ is a homogeneous polynomial of degree three in $Z_1, \ldots, Z_n$, unbiased for γ_1, and, under normality, of minimal variance among all these estimators.

b) The Aitken estimate for γ_1 in the derived model

$$Z^{*3} \sim (M^{*3} 1_n \sigma^3 \gamma_1 ; \sigma^6 S_G)$$

is given by

$$\hat{g}_1 := (\sigma^3 1_n' M^{*3} S_G^+ M^{*3} 1_n)^{-1} 1_n' M^{*3} S_G^+ Z^{*3}.$$

$\hat{g}_1$ is a homogeneous polynomial of degree three involving pure cubes $Z_1^3, \ldots, Z_n^3$ only, unbiased for γ_1, and, under normality, of minimal variance among all these estimators.

c) The Least Squares Estimator in the model generated by $Z^{\otimes 3}$ is

$$\overline{g_1} := (\sigma^3 1_n' M^{*3} 1_n)^{-1} 1_n' Z^{*3}.$$

$\overline{g_1}$ is a homogeneous polynomial of degree three involving pure cubes $Z_1^3, \ldots, Z_n^3$ only, unbiased for γ_1, and of minimal norm among all unbiased homogeneous polynomials of degree three in $Z_1, \ldots, Z_n$.

Under normality, these estimates are related as stated in d) – g):

d) $\operatorname{Var} \hat{\gamma}_1 \leqslant \operatorname{Var} \hat{g}_1 \leqslant \operatorname{Var} \overline{g_1} \leqslant \frac{1}{\nu} [6 (\nu^2/n^2 - p/n)^{-1} + 9 (\nu^2/n^2 - p/n)^{-2}]$, where $\nu = n - p = \operatorname{rank} M$.

e) $\hat{g}_1$ is of minimal variance in Γ_1 if and only if

$$\pi_s S_\Gamma H_3' S_G^+ M^{*3} 1_n = \rho \sum_{i=1}^{n} (Me_i)^{\otimes 3}, \ \rho \in \mathbf{R}.$$

f) $\overline{g_1}$ is of minimal variance in G_1 if and only if

$$(\operatorname{Diag} M) Mm = \rho M^{*3} 1_n, \ \rho = m'Mm / 1_n' M^{*3} 1_n.$$

g) $\overline{g_1}$ is of minimal variance in Γ_1 if and only if $\pi_s \cdot (Mm) \otimes (\operatorname{vec} M) = \rho \sum_{i=1}^{n} (Me_i)^{\otimes 3}$, $\rho = m'Mm / 1_n' M^{*3} 1_n$.

Proof: In a general linear model $Y \sim (X\beta;\ \sigma^2 V)$, the Aitken estimate $(X'V^+X)^+ X'V^+Y$ for β is of mimimun variance iff range $X \subset$ range V [*Zyskind*, p. 658]. Since for non-negative-definite matrices the range of a sum equals the sum of the ranges, this condition is easily verified for a) and b). c) uses the representation X^+Y of the Least Squares Estimator. The first two inequalities in d) are immediate, the last one uses Lemma 2.4. e) – g) follow since, in general, $\hat{c}'Y$ is *Blue* of its expectation iff $V\hat{c}$ lies in the range of X [*Zyskind*, p. 653]. For case g), e.g., this gives $\pi_s S_\Gamma \Sigma e_i^{\otimes 3} = \rho \Sigma (Me_i)^{\otimes 3}$ for some real number ρ. Its essential term is $\pi_s \cdot \Sigma (Me_i) \otimes \mathrm{vec}\, M \cdot M_{ii} = \pi_s \cdot$ $\cdot (Mm) \otimes \mathrm{vec}\, M$. In each of the assertions e) – g), the particular value of ρ can be computed by suitable premultiplication. □

Examples: Anscombe [1961, pp. 14f.] contents himself with estimating the third moment $\mu_3 = \sigma^3\gamma_1$ for which problem Theorem 1 holds true using the estimates $\sigma^3\hat{\gamma}_1$, $\sigma^3\hat{g}_1$, and $\sigma^3\overline{g_1}$, respectively; the classes Γ_1 and G_1 remain the same. *Anscombe*'s $\sigma^3 g_1$ coincides with our least squares version $\sigma^3\overline{g_1}$.

Firstly, for $Y \sim ((0, 1, 2)'\mu;\ \sigma^2 I_3)$, Theorem 1 yields $\sigma^3\overline{g_1} = \frac{125}{174}(Z_1^3 + Z_2^3 + Z_3^3)$; this is of minimum variance in G_1, but not in Γ_1 [*Anscombe*, p. 14].

Secondly, for $Y \sim ((0, 1, 2, 3)'\mu;\ \sigma^2 I_4)$, the variance of $\sigma^3\overline{g_1}$ is not minimal in G_1. For $\sigma^3\hat{g}_1$, one obtains $\sigma^3\hat{g}_1 = 0.365\, Z_1^3 + 0.573\, Z_2^3 + 1.015\, Z_3^3 + 2.404\, Z_4^3$, and $\mathrm{Var}\, \hat{g}_1 = 5.48 < 6.59 = \mathrm{Var}\, \overline{g_1}$ [*Anscombe*, p. 15].

Thirdly, consider *Anscombe's design-condition I* [1961, p. 3] of a common over-all mean component: $1_n \in \mathrm{range}\, X$, and *design-condition II* of equal variances of the residuals $Z_1, \ldots, Z_n$: $M_{11} = \cdots = M_{nn}$. Under these conditions $Mm = 0$, and Theorem 1 g) ensures that $\sigma^3\overline{g_1}$ is of minimal variance in the big class Γ_1, and not only in its subclass G_1, as found by *Anscombe* [1961, p. 15]. □

4. Kurtosis

The following discussion parallels that of Section 3. First the dispersion matrix of $Z^{\otimes 4}$ is computed.

Lemma 4.1: For every quasinormal classical linear model the dispersion matrices of $Z^{\otimes 4}$ and Z^{*4} are given by $\sigma^8 \pi_s A_\Gamma \pi_s$ and $\sigma^8 A_G$, respectively, where

$$A_\Gamma := 24\, M^{\otimes 4} + 72\, M \otimes M \otimes (\mathrm{vec}\, M \cdot \mathrm{vec}' M)\ .$$

$$A_G := 24\, M^{*4} + 72\, (\mathrm{Diag}\, M)\, M^{*2}\ (\mathrm{Diag}\, M)\ .$$

Proof: Define $A := 24\, I_n^{\otimes 4} + 72\, I_n^{\otimes 2} \otimes (\mathrm{vec}\, I_n \cdot \mathrm{vec}'\, I_n) + 9\, (\mathrm{vec}\, I_n \cdot \mathrm{vec}'\, I_n)^{\otimes 2}$. As in Lemma 3.1 first prove $EU^{\otimes 4}\, (U^{\otimes 4})' = \pi_s A \pi_s$; using $EU^{\otimes 4} = 3\, (\mathrm{vec}\, I_n)^{\otimes 2}$ one then finds $\mathcal{D}U^{\otimes 4}$, $\mathcal{D}Z^{\otimes 4}$, and $\mathcal{D}Z^{*4}$ as asserted. □

Note that one may now obtain equation (33) in *Anscombe* [1961].

Theorem 2 is our main result on estimating the coefficient of kurtosis γ_2. As in Theorem 1, it is assumed that σ^2 is known.

Theorem 2: Assume a classical linear model $Y \sim (X\beta; \sigma^2 I_n)$.

a) The Aitken estimate for γ_2 in the derived model

$$Z^{\otimes 4} - 3\,\sigma^4 \pi_s \cdot (\operatorname{vec} M)^{\otimes 2} \sim \Big(\sum_{i=1}^{n} (Me_i)^{\otimes 4} \cdot \sigma^4 \gamma_2 ; \sigma^8 \pi_s A_\Gamma \pi_s\Big)$$

is given by

$$\hat{\gamma}_2 := \frac{\sigma^{-4} \sum_{i=1}^{n} e_i'^{\otimes 4} (\pi_s A_\Gamma \pi_s)^+ Z^{\otimes 4} - 3 \cdot \sum_{i=1}^{n} e_i'^{\otimes 4} (\pi_s A_\Gamma \pi_s)^+ (\operatorname{vec} M)^{\otimes 2}}{\sum_{i,j=1}^{n} e_i'^{\otimes 4} (\pi_s A_\Gamma \pi_s)^+ e_j^{\otimes 4}}.$$

$\hat{\gamma}_2$ is a constant plus a homogeneous polynomial of degree four in $Z_1, \ldots, Z_n$, unbiased for γ_2, and, under normality, of minimal variance among all these estimators.

b) The Aitken estimate for γ_2 in the derived model

$$Z^{*4} - 3\,\sigma^4 m^{*2} \sim (M^{*4} 1_n \sigma^4 \gamma_2 ; \sigma^8 A_G)$$

is given by

$$\hat{g}_2 := \frac{\sigma^{-4} 1_n' M^{*4} A_G^+ Z^{*4} - 3 \cdot 1_n' M^{*4} A_G^+ m^{*2}}{1_n' M^{*4} A_G^+ M^{*4} 1_n}.$$

$\hat{g}_2$ is a constant plus a homogeneous polynomial of degree four involving pure fourth powers $Z_1^4, \ldots, Z_n^4$ only, unbiased for γ_2, and, under normality, of minimal variance among all these estimators.

c) The Least Squares Estimator in the model as in a) is

$$\overline{g_2} := (1_n' M^{*4} 1_n)^{-1} (\sigma^{-4} 1_n' Z^{*4} - 3 \cdot m'm).$$

$\overline{g_2}$ is a constant plus a homogeneous polynomial of degree four involving pure fourth powers $Z_1^4, \ldots, Z_n^4$ only, unbiased for γ_2, and of minimal norm among all unbiased constants plus homogeneous polynomials of degree four in $Z_1, \ldots, Z_n$.

Under normality, these estimates are related as stated in d) – g):

d) $\operatorname{Var} \hat{\gamma}_2 \leqslant \operatorname{Var} \hat{g}_2 \leqslant \operatorname{Var} \overline{g_2} \leqslant \frac{1}{\nu} [24 (n^3/\nu^3) + 72 (n^6/\nu^6)]$, where $\nu = \operatorname{rank} M$.

e) $\hat{g}_2$ is of minimal variance in Γ_2 if and only if

$$\pi_s A_\Gamma H_4' A_G^+ M^{*4} 1_n = \rho \sum_{i=1}^{n} (Me_i)^{\otimes 4}, \ \rho \in \mathbf{R}.$$

f) $\overline{g_2}$ is of minimal variance in G_2 if and only if

$$(\operatorname{Diag} M)\, M^{*2}\, m = \rho\, M^{*4}\, 1_n, \quad \rho = m'M^{*2}m / 1_n'\, M^{*4}\, 1_n.$$

g) $\overline{g_2}$ is of minimal variance in Γ_2 if and only if

$$\pi_s \cdot (\operatorname{vec} M) \otimes (\operatorname{vec} M\, (\operatorname{Diag} M)\, M) = \rho \sum_{i=1}^{n} (Me_i)^{\otimes 4}, \quad \rho \text{ as in (f).} \quad \square$$

Examples: For the first example of Section 3, $\overline{g_2} = \sigma^{-4}\, \frac{625}{914} (Z_1^4 + Z_2^4 + Z_3^4) - \frac{3150}{914}$.
This is of minimal variance in G_2, but not in Γ_2.

In the second case, $\overline{g_2}$ does not minimize the variance in G_2. For $\hat{g}_2$ we obtained $\hat{g}_2 = \sigma^{-4}\, (0.4843\, Z_1^4 + 0.4841\, Z_2^4 + 0.5956\, Z_3^4 - 0.3997\, Z_4^4) - 3.4639$, and $\operatorname{Var} \hat{g}_2 = 46.49 < 47.17 = \operatorname{Var} \overline{g_2}$.

Thirdly, let us introduce design-condition III [cf., *Anscombe*, pp. 6,20] of up to permutations equal covariances of the residuals, or precisely: the rows of M are equal up to permutations of their entries. Then design-conditions II and III force $\overline{g_2}$ to have minimal variance in G_2: in Theorem 2 f), $(\operatorname{Diag} M)\, M^{*2} m = M_{11}^2\, M^{*2}\, 1_n$ is a multiple of 1_n, and the same is true of $M^{*4}\, 1_n$. This is an example with a nontrivial ρ.

Fourthly, we replace in $\overline{g_2}$ the generally unknown σ^{-4} by an estimate, and compare our $\tilde{g}_2$, obtained along these lines, with *Anscombe*'s g_2. As before, define $\nu := \operatorname{rank} M = n - \operatorname{rank} X;\ d := \sum_{i,j=1}^{n} M_{ij}^4$. Estimate σ^2 by $s^2 = \nu^{-1} \sum_{i=1}^{n} Z_i^2$, as usual.
Replacing σ^2 by s^2 generates a bias even under normality:
$E s^{-4}\, 1_n'\, Z^{*4} - 3\, m'm = -6\, (\nu + 2)^{-1} m'm$. Correcting for this bias by changing the constant yields

$$\tilde{g}_2 := \frac{1}{d}\, (s^{-4} \sum Z_i^4 - 3\, (\nu + 2)^{-1} \nu m'm).$$

Anscombe [1961, p. 7] uses a different divisor:

$$g_2 = \frac{d}{D} \tilde{g}_2,\ D := d - 3\, [\nu\, (\nu + 2)]^{-1}\, [\sum_{i=1}^{n} M_{ii}^2]^2.$$

Since $d/D > 1$, the variance of g_2 always exceeds that of $\tilde{g}_2$. Under normality, both estimators are unbiased for $\gamma_2 = 0$. The asymptotic properties are the same: if (*) $\sup \{\operatorname{rank} X_n \mid n = 1, 2, \ldots\} \leqslant p$, then

$$0 \leqslant 1 - D/d = 3\, (1_n'\, m^{*2})^2 \,/\, [\nu\, (\nu + 2) \sum M_{ij}^4]$$

$$\leqslant 3n \sum M_{ii}^4 \,/\, [\nu\, (\nu + 2)\, (\sum M_{ii}^4 + \sum_{i \neq j} M_{ij}^4)]$$

$$\leqslant 3n / [(n - p)\, (n - p + 2)] \to 0$$

as $n \to \infty$. For the model $Y \sim (1_n \mu; \sigma^2 I_n)$, *Anscombe*'s g_2 specializes to Fisher's kurtosis statistic [cf., *Cramèr,* eq. (29.3.8)], whereas $\widetilde{g_2}$ does not.

Finally note that under condition (*) all variances in Theorems 1 d) and 2 d) tend to zero, whence the proposed estimators are consistent. □

We remark that for the estimation of the kurtosis one may utilize yet another derived model, namely,

$$Z^{\otimes 4} \sim ([\sum_{i=1}^{n} (Me_i)^{\otimes 4} : 3\pi_s (\operatorname{vec} M)^{\otimes 2}] \begin{pmatrix} \sigma^4 \gamma_2 \\ \sigma^4 \end{pmatrix}; \sigma^8 \pi_s A_\Gamma \pi_s).$$

The approach chosen above reveals more clearly the analogy of estimating skewness and kurtosis, and also conforms to a greater extent with *Anscombe*'s paper [1961].

Acknowledgements

I would like to thank Professor *H. Witting* for valuable discussions throughout the course of this work. I am also grateful to the referees for helpful comments. This work originated with my doctoral thesis and was completed while I was on a postdoctoral fellowship of the Deutsche Forschungsgemeinschaft, Bonn, with the Department of Statistics, Stanford Univeristy. A preliminary version of this paper appeared as Technical Report No. 103 NSF Grant MPS74-21416 at Stanford.

References

Anscombe, F.J.: Examination of residuals. Proc. Fourth Berkeley Symp. Math. Statist. Prob. **1**, 1961, 1–36.

Cramèr, H.: Mathematical Methods of Statistics. Princeton 1946.

Greub, W.H.: Multilinear Algebra. Berlin 1967.

Pukelsheim, F.: On Hsu's model in regression analysis. Math. Operationsforsch. Statist. Ser. Statist. **8**, 1977, 323–331.

Rao, C.R.: Linear Statistical Inference and Its Applications. Second Edition. New York 1973.

Schmetterer, L.: Introduction to Mathematical Statistcs. New York 1974.

Styan, G.P.H.: Hadamard products and multivariate statistical analysis. Linear Algebra and Appl. **6**, 1973, 217–240.

Zyskind, G.: Error structures, projections, and conditional inverses in linear model theory. A Survey of Statistical Design and Linear Models. Ed. by J.N. Srivastava, Amsterdam 1975, 647–663.

Received October 26, 1977
(revised version September 6, 1978)