DATA TRANSFORMATIONS AND THE LINEAR MODEL

By D. A. S. Fraser

University of Toronto

0. Summary. The familiar application of the normal linear model involves a response variable that is assumed normally distributed with constant variance and with location linear in a vector parameter. In other applications a response variable may occur in a form that suppresses an underlying normal linear structure. Sometimes in these applications the context may suggest a logarithmic or square root transformation which reveals the normal linear form. Box and Cox (1964) consider a parametric class of transformations on the response variable and derive a method for estimating the class parameter based on Bayesian and likelihood techniques.

In this paper a more comprehensive statistical model is proposed; it is a revision of the structural model (Fraser 1966). It gives stronger inference statements in the context for the linear model. It can handle normal or nonnormal error. And in the context for the transformed linear model it gives directly a method for estimating the class parameter. This estimation avoids an approximation in the Bayesian prior and it includes additional sensitivity to the data.

1. The linear structural model. This section develops the linear structural model and records the formulas for the general structural model.

(1.1) *The Model.* Consider a response variable y based on a process operating under stable conditions. Variation in a response is usually attributable to a variety of sources: variation in the material being processed; variation in the conditions of the process; variation in the internal operation of the process. These sources of variation form the *internal error* of the system.

Now suppose that the external conditions of the process are controlled and observations are randomly sequenced against any remaining identifiable sources of variation. This may provide the basis on which the internal error has stability and on which the effect of this error *in* the response has known form with independence between observations. The production of the internal error is the *random process* of the system and the effect of this in the response is the *error variable* of the system. These characteristics require a scale of measurement; they do not require a unit of measurement on this scale, and they do not require an origin. Let e be the error variable in some arbitrary units and let $f(e)\,de$ be its probability element on the real line.

The medial or general level of the response is the quantity typically of interest and it depends in general on the variables being controlled. This quantity can have numerical definition by introducing a unit of measurement, and by introducing an origin of measurement for any chosen set of conditions. Suppose experience with similar systems gives grounds for assuming that the medial level

Received 22 December 1966; revised 30 April 1967.

is linear in some combination of variables constructed from the variables being controlled.

Now consider a succession of n response observations corresponding to a succession of levels chosen for the controllable variables. Let y_i be the ith observation on the response, and $v_{1i}, \cdots, v_{ri}$ be the corresponding values of the constructed variables. And let σ be the response scaling of the error variable and $\beta_1, \cdots, \beta_r$ be the quantities that give the general levels of the response from the levels of the constructed variables. This gives the linear structural model

$$\prod_1^n f(e_i) \prod_1^n de_i$$

$$\begin{aligned} y_1 &= \beta_1 v_{11} + \cdots + \beta_r v_{r1} + \sigma e_1 \\ &\vdots \\ y_n &= \beta_1 v_{1n} + \cdots + \beta_r v_{rn} + \sigma e_n . \end{aligned}$$

The model has two parts: an *error distribution* $\prod f(e_i) \prod de_i$ describing the effect of multiple operation of the internal error (the e's are variables); and a composite *structural equation* in which a *realized* vector $\mathbf{e} = (e_1, \cdots, e_n)'$ from the error distribution determines the relationship between the *known* observations $\mathbf{x} = (x_1, \cdots, x_n)'$ and the *unknown* physical quantities $\beta_1, \cdots, \beta_r, \sigma$ (the e's are unknown constants).

(1.2) *The transformations.* The structural equation presents the response vectors as a transformation of the error vector; this can be expressed conveniently in matrix notation as

$$\begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \\ y_1 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \\ & & \cdot & & \vdots \\ & & & \cdot & \\ 0 & & & 1 & 0 \\ \beta_1 & \cdots & & \beta_r & \sigma \end{bmatrix} \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \\ e_1 & \cdots & e_n \end{bmatrix}$$

or as $Y = \theta E$ where

$$Y = (\mathbf{v}_1' \cdots \mathbf{v}_r' \mathbf{y}')', \qquad E = (\mathbf{v}_1' \cdots \mathbf{v}_r' \mathbf{e}')'.$$

Consider the effect of the *regression group*

$$G = \left\{ g = \begin{bmatrix} 1 & & & 0 \\ & \cdot & & \vdots \\ & & \cdot & \\ 0 & & 1 & 0 \\ a_1 & \cdots & a_r & c \end{bmatrix} : -\infty < a_i < \infty, 0 < c < \infty \right\}$$

of such transformations on Euclidean n-space. A point Y is carried by the elements of G into the *orbit* $GY = \{gY : g \in G\}$. The bottom row of Y gives the vector point in n-space; the remaining rows are surplus and allow group multiplication to be matrix multiplication. The orbits partition n-space: assume the levels of controllable variables were chosen to avoid the triviality of linear de-

pendence among the $\mathbf{v}$'s; and delete the linear subspace of the $\mathbf{v}$'s as a trivial orbit. The orbits are then disjoint $r + 1$ dimensional half-spaces.

(1.3) *Transformation variables.* The position of a point Y on its orbit can be described by a transformation variable:

DEFINITION. $[Y]$ is a transformation variable if $[Y]$ takes values in G and $[gY] = g[Y]$ for all g, Y.

A transformation gives a *reference point* on each orbit: $D(Y) = [Y]^{-1}Y$. The reference points index the orbits and the transformation variable gives *position* on an orbit: $Y = [Y]D(Y)$. Two transformation variables differ on any orbit by right multiplication by a group element.

A transformation variable can be obtained by regression analysis:

$$[Y] = \begin{bmatrix} 0 & & 0 & 0 \\ & \ddots & & \vdots \\ 1 & & 1 & 0 \\ b_1(Y) & \cdots & b_r(Y) & s(Y) \end{bmatrix}$$

where $b_1, \cdots, b_r$, and s are respectively the regression coefficients of $\mathbf{y}$ on $\mathbf{v}_1, \cdots, \mathbf{v}_r$, and the residual length. The corresponding reference point is

$$D(Y) = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \\ v_{r1} & \cdots & v_{rn} \\ d_1 & \cdots & d_n \end{bmatrix}$$

where

$$\mathbf{d}(Y) = s^{-1}(Y)[\mathbf{y} - b_1(Y)\mathbf{v}_1 - \cdots - b_r(Y)\mathbf{v}_r]$$

is the unit residual vector. And as noted by W. Keith Hastings a transformation variable can also be obtained from linear programming algorithms by finding coefficients that minimize the sum of absolute deviations; this avoids the quadratic calculations of least squares.

The linear structural model can now be written

$$f(E)\,dE$$

$$GY = GE, \qquad [Y] = \theta[E].$$

The model has two parts: an *error distribution* $f(E)\,dE = \prod f(e_i) \prod de_i$ (E is a variable); and a composite structural equation in which a *realized* error E from the error distribution gives the relationship between the known observation Y and the unknown quantity θ (E is an unknown constant).

More generally, let E be a variable on a space Y with probability element $f(E)\,dE$ based on a fixed measure, and let G be a group of transformations that is unitary on Y and preserves the σ-algebra of the measure. With these substitutions the preceding formulas describe the *general structural model.*

(1.4) *Homogeneity.* Consider a transformation g in G,

$$Y^* = gY, \qquad \theta^* = g\theta,$$

and view the transformation as providing new coordinates for the observation Y and the quantity θ. A transformation g changes the response unit, and it changes the response origins differentially among the individual observations.

Consider the effect on the model. The structural equation $Y = \theta E$ becomes $Y^* = \theta^* E$. Thus the model

$$f(E)\ dE$$
$$Y = \theta E$$

becomes

$$f(E)\ dE$$
$$Y^* = \theta^* E.$$

The physical problem is untouched by the transformation. The model reflects this and has the *same form* point for point under the change of coordinates. The model is said to be *homogeneous under the regression group.*

(1.5) *Probabilities for constants.* Consider probability statements for unknown constants.

Suppose 2 cards are dealt face down on a table from a well shuffled deck of playing cards. The designations on the 2 cards are unknown constants. An observer can make a probability assertion

$$\Pr\{2 \text{ diamonds}\} = (13/52)(12/51),$$

based on the *random process* that generated the unknown constants.

Now suppose 2 further cards are dealt face down and suppose the observer examines them and finds the first is a diamond and the second a non diamond. The observer can make a revised probability assertion

$$\Pr\{2 \text{ diamonds}\} = (13/52)(12/51)(11/50)(39/49)/(13/52)(39/51)$$
$$= (13/50)(12/49)(11/13)$$

based on the *random process as conditioned by the observed event.*

Alternatively, suppose the second pair of cards is kept face down and passed to a participant in an adjacent room. And suppose the participant reports the item of information "There's a diamond here". The observer might assert

$$\Pr\{2 \text{ diamonds}\} = (13/52)(12/51)(11/50)/(13/52) = (13/51)(12/50)(11/13)$$

if he thought the participant had examined only the first card. Or he might assert

$$\Pr\{2 \text{ diamonds}\} = (13/52)(12/51)(2(11/50)(39/49))/2(13/52)(39/51)$$
$$= (13/50)(12/49)(11/13),$$

if he thought the participant had examined both and would have reported 2 diamonds if there were 2 diamonds. The two assertions are contradictory.

In this alternative situation a correct assertion giving a value to Pr{2 diamonds} *cannot* be made. The item of information "There's a diamond here", is not an *event* but has the form of a *deduction* from some unknown event.

The example illustrates sufficient conditions for making probability assertions about unknown constants: (i) *The constants were generated as realized values from a random process with known probability characteristics.* (ii) *The only other information concerning the unknown constants has the form of an event for the random process that generated the constants.*

(1.6) *Reduction.* Consider an application of the linear structural model. And suppose there is no outside information concerning θ. This could arise minimally if the system is being examined in isolation to see what information it alone supplies concerning θ.

Consider the information in the structural equation

$$GY = GE, \qquad [Y] = \theta[E]$$

concerning the unknown realized error E. The orbit of E is known: $GY = GE$. And it is known in the form of an event based on the variable GE. The position of E on its orbit is *not* known:

$$[E] = \theta^{-1}[Y] = g[Y].$$

The unknown position $[E]$ is represented as an unknown transformation g applied to a position value $[Y]$. If the known position were different, say $[Y^*] = h[Y]$, then the unknown position $[E]$ would be represented as $[E] = gh[Y] = g^*[Y]$ where g^* is also an unknown transformation in the group (this uses the homogeneity of the model). Thus different values for the transformation $[Y]$ would provide the *same* description of $[E]$. There is thus no information concerning the position $[E]$.

The only information concerning the unknown E has the form $GE = GY$, an event for the random process that generated E. It follows that *exact probability statements can be made concerning the unknown* E; these are based on *the conditional distribution of the error variable* E *given the orbit* $GE = GY$.

(1.7) *The reduced model.* The conditional distribution of the error variable E given the orbit GE is easily derived using invariant differentials.

On the space R^n the transformation $Y \to gY$ has Jacobian $|\partial gY/\partial Y| = c^n$. A scale variable is $s(Y)$; accordingly, *an* invariant differential is

$$dm(Y) = dY/s^n(Y).$$

On the group G the left transformation $h \to gh$ has Jacobian $|\partial gh/\partial h| = c^{r+1}$. Accordingly, *the* invariant differential is

$$d\mu(g) = dg/c^{r+1}$$

(up to a constant multiplier). Similarly the right transformation has invariant differential

$$d\nu(g) = dg/c.$$

These are related by

$$d\mu(g) = \Delta(g)\, d\nu(g) = \Delta(g)\, d\mu(g^{-1}), \qquad \Delta(g) = c^{-r}.$$

The probability element along a neighbourhood of orbits is

$$f(E)\, dE = \bar{f}(E) \cdot dm(Y) = a(D)\bar{f}([E]D)\, d\mu([E]).$$

The conditional probability element for $[E]$ given $[E]^{-1}E = D$ is obtained by normalization:

$$\begin{aligned} g([E]:D)\, d[E] &= \bar{g}([E]:D)\, d\mu([E]) \\ &= k(D)\bar{f}([E]D)\, d\mu([E]) \\ &= k(D)f([E]D)s^{n-r-1}(E)\, d[E]. \end{aligned}$$

The information in the linear structural model produces the reduced model

$$\begin{gathered} \bar{g}([E]:D)\, d\mu([E]) \\ [Y] = \theta[E]. \end{gathered}$$

The reduced model has an *error probability distribution* $\bar{g}([E]:D)\, d\mu([E])$,

$$k(D)f([E]D)s^{n-r-1}(E)\, d[E],$$

which provides exact probability statements concerning the unknown constant $[E]$ in the structural equation; and it has a *structural equation* $[Y] = \theta[E]$,

$$\begin{aligned} b_1(Y) &= \beta_1 + \sigma b_1(E) \\ &\vdots \qquad\qquad \vdots \\ b_r(Y) &= \beta_r + \sigma b_r(E) \\ s(Y) &= \sigma s(E), \end{aligned}$$

in which the unknown $[E]$ determines the relationship between the known position $[Y]$ and the unknown θ.

The formulas in this section, except for occasional specialization are those of the general structural model as described in Section (1.3).

For the case of a normal error variable

$$f(E)\, dE = (2\pi)^{-n/2} \exp\{-\tfrac{1}{2}\textstyle\sum e_i^2\} \prod_1^n de_i$$

the conditional distribution takes the form

$$\begin{aligned} k(D)f([E]D)s^{n-r-1}\, d[E] &= k(D) \exp\{-\tfrac{1}{2}\textstyle\sum (s(E)\, d_i + \sum b(E)v_i)^2\} s^{n-r-1}\, d[E] \\ &= (|A|^{\frac{1}{2}}/(2\pi)^{r/2}) \exp\{-\tfrac{1}{2}\textstyle\sum b_t b_u a_{tu}\}\, d\mathbf{b}. \\ &\quad (A_{n-r}/(2\pi)^{(n-r)/2} \exp\{-\tfrac{1}{2}s^2\} s^{n-r-1}\, ds \end{aligned}$$

where $A = (a_{tu})$ is the inner product matrix for the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_r$ and $A_f^{\#} = 2\pi^{f/2}/\Gamma(f/2)$ is the area of a unit sphere in f-space. The conditional distribution does not depend on D for this case of normal error.

(1.8) *Inference.* Suppose a value $\beta_r = \beta_{r0}$ has been suggested by some outside source. The hypothesis: $\beta_r = \beta_{r0}$ leads to a further characteristic of the error E:

$$b_r(E)/s(E) = (b_r(Y) - \beta_{r0})/s(Y).$$

This *value* can be compared with the distribution of the *variable* $b_r(E)/s(E)$ based on the error distribution of the reduced model, and the hypothesis assessed accordingly.

Suppose that general information concerning the value of θ is wanted. The reduced model gives a probability distribution describing the unknown value $[E]$. And it gives a structural equation: $[Y] = \theta[E]$, $[E] = \theta^{-1}[Y]$. For each possible value for $[E]$ there corresponds a possible value for θ ($[Y]$ is fixed and known). The error distribution describing the unknown $[E]$ is thus *ipso facto* a distribution for θ, *the structural distribution for* θ:

$$\begin{aligned}\bar{g}(\theta^{-1}[Y]:D)\, d\mu(\theta^{-1}[Y]) &= \bar{g}(\theta^{-1}[Y]:D)\Delta(\theta^{-1}[Y])\, d\mu(\theta) \\ &= k(D)\bar{f}(\theta^{-1}Y)\cdot\Delta(\theta^{-1}[Y])\, d\mu(\theta) \\ &= k(D)f(\theta^{-1}Y)(s(Y)/\sigma)^{n}\cdot(s(Y)/\sigma)^{-r}\, d\beta d\sigma/\sigma^{r+1}.\end{aligned}$$

For the normal error case the structural distribution is

$$\begin{aligned}&(|A|^{\frac{1}{2}}/(2\pi)^{r/2}) \exp\{-(1/2\sigma^2)\textstyle\sum (b_t(Y) - \beta_t)(\beta_u(Y) - \beta_u)a_{tu}\}\, d\beta \\ &\qquad\cdot(A_{n-r}/(2\pi)^{(n-r)/2}) \exp\{-\tfrac{1}{2}(s(Y)/\sigma)^2\}\ (s(Y)/\sigma)^{n-r}\, d\sigma/\sigma^{r+1}.\end{aligned}$$

2. The transformed linear structural model. Consider a response variable y and a class of transformations

$$y^{(\lambda)} = l(y:\lambda).$$

And suppose that for some λ the transformed response can be described by a linear structural model:

$$f(E)\, dE$$

$$Y^{(\lambda)} = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \\ y_1^{(\lambda)} & \cdots & y_n^{(\lambda)} \end{bmatrix} \theta = \begin{bmatrix} v_{11} & & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \\ e_1 & \cdots & e_n \end{bmatrix} = \theta E.$$

Box and Cox (1964) suggest two examples for the transformation $l(y:\lambda)$:

$$\begin{aligned} l(y:\lambda) &= y^{\lambda}, \qquad \lambda \neq 0, \\ &= \ln y, \qquad \lambda = 0, \end{aligned}$$

and

$$\begin{aligned} l(y:\lambda) &= (y + \lambda_2)^{\lambda_1}, \qquad \lambda_1 \neq 0, \\ &= \ln(y + \lambda_2), \qquad \lambda_1 = 0. \end{aligned}$$

The first and simpler example gives the three common-transformations, the square-root, the reciprocal, and the logarithmic. For ease of notation, the brackets on the *index* λ will be deleted in the remainder of this section: $y^{(\lambda)} \to y^{\lambda}$.

First consider inference concerning θ assuming that the value of the class parameter λ is known.

The reduced model is

$$k(D^{\lambda})f([E]D^{\lambda})s^{n-r-1}\,d[E]$$
$$[Y^{\lambda}] = \theta[E].$$

The structural distribution for θ is

$$k(D^{\lambda})f(\theta^{-1}Y^{\lambda})(s(Y^{\lambda})/\sigma)^{n-r}\,d\beta d\sigma/\sigma^{r+1}$$

and for normal error is

$$(|A|^{\frac{1}{2}}/(2\pi)^{r/2})\exp\{-(1/2\sigma^2)\textstyle\sum(b_t(Y^{\lambda})-\beta_t)(b_u(Y^{\lambda})-\beta_u)a_{tu}\}\,d\beta$$
$$\cdot(A_{n-r}/(2\pi)^{(n-r)/2})\exp\{-\tfrac{1}{2}(s(Y^{\lambda})/\sigma)^2\}(s(Y^{\lambda})/\sigma)^{n-r}\,d\sigma/\sigma^{r+1}$$

Now consider inference concerning λ. The structural equation is

$$GE = GY^{\lambda}, \qquad [Y^{\lambda}] = \theta[E].$$

With θ unknown the second component gives no information concerning λ. Hence consider the first component $GE = GY^{\lambda}$ and evaluate it in terms of the classical model.

The probability element for $[Y^{\lambda}]$ given D^{λ} is

$$k(D^{\lambda})f(\theta^{-1}Y^{\lambda})(s(Y^{\lambda})/\sigma)^{n}\,d[Y^{\lambda}]/s^{r+1}(Y^{\lambda}).$$

The transformation $Y^{\lambda} = l(y:\lambda)$ produces a change in metric; the relationship between differentials is

$$(dy_1^{\lambda}, \cdots, dy_n^{\lambda}) = (dy_1, \cdots, dy_n)J(\mathbf{y}:\lambda),$$
$$(dy_1, \cdots, dy_n) = (dy_1^{\lambda}, \cdots, dy_n^{\lambda})J^{-1}(\mathbf{y}:\lambda)$$

where

$$J(\mathbf{y}:\lambda) = \begin{bmatrix} dy_1^{\lambda}/dy_1 & & 0 \\ & \ddots & \\ 0 & & dy_n^{\lambda}/dy_n \end{bmatrix}$$

is the Jacobian matrix of the transformation. The transformation $Y^{\lambda} = [Y^{\lambda}]D^{\lambda}$ to position $[Y^{\lambda}]$ on the orbit D^{λ} also gives a change in metric; the relationship between differentials is

$$(dy_1^{\lambda}, \cdots, dy_n^{\lambda}) = (db_1, \cdots, db_r, ds)D^{\lambda}.$$

For the composite transformation the relationship between differentials is then

$$(dy_1, \cdots, dy_n) = (db_1, \cdots, db_r, ds)D^\lambda J^{-1}(\mathbf{y}:\lambda).$$

For fixed D^λ an $r+1$ dimensional subspace of R^n is generated by the composite transformation. Let v^λ be Euclidean volume in this subspace; then

$$dv^\lambda = |D^\lambda J^{-2}(\mathbf{y}:\lambda)D'^\lambda|^{\frac{1}{2}}\, d\mathbf{b}\, ds.$$

The probability element of $[Y^\lambda]$ given D^λ can then be expressed in terms of the element v^λ:

$$k(D^\lambda)f(\theta^{-1}[Y^\lambda])[(s(Y^\lambda))^{n-r-1}/\sigma^n]\, dv^\lambda/|D^\lambda J^{-2}(\mathbf{y}:\lambda)D'^\lambda|^{\frac{1}{2}}.$$

The probability element for the response Y is

$$f(\theta^{-1}Y^\lambda)(1/\sigma^n)\, dY^\lambda = f(\theta^{-1}Y^\lambda)(1/\sigma^n)\, dY/|J(\mathbf{y}:\lambda)|^{-1}.$$

The probability element for the response Y divided by the *conditional* element for the position $[Y^\lambda]$ gives the *marginal* probability element for the orbital variable D^λ:

$$(1/s^{n-r-1}(Y^\lambda))[|D^\lambda J^{-2}(\mathbf{y}:\lambda)D'^\lambda|^{\frac{1}{2}}/k(D^\lambda)\, |J(\mathbf{y}:\lambda)|^{-1}]\cdot dY/dv^\lambda.$$

In the case of normal error the marginal element for D^λ is

$$\begin{aligned}(1/s^{n-r-1}(Y^\lambda))&[|D^\lambda J^{-2}(\mathbf{y}:\lambda)D'^\lambda|^{\frac{1}{2}}/A_{n-r}\, |A|^{\frac{1}{2}}\, |J(\mathbf{y}:\lambda)|^{-1}]\, dY/dv^\lambda \\ &= (1/s^{n-r-1}(Y^\lambda))\cdot[|D^\lambda J^{-2}(\mathbf{y}:\lambda)D'^\lambda|^{\frac{1}{2}}/A_{n-r}\, |D^\lambda D'^\lambda|^{\frac{1}{2}}\, |J(\mathbf{y}:\lambda)|^{-1}]\, dY/dv^\lambda.\end{aligned}$$

The *marginal likelihood* function for λ is then

$$L(\lambda:D^\lambda) = (1/s^{n-r-1}(Y^\lambda))\cdot[|D^\lambda J^{-2}(\mathbf{y}:\lambda)D'^\lambda|^{\frac{1}{2}}/k(D^\lambda)\, |J(\mathbf{y}:\lambda)|^{-1}]$$

and in the normal case is

$$L(\lambda:D^\lambda) = (1/s^{n-r-1}(Y^\lambda))\cdot|D^\lambda J^{-2}(\mathbf{y}:\lambda)D'^\lambda|^{\frac{1}{2}}/|D^\lambda D'^\lambda|^{\frac{1}{2}}\, |J(\mathbf{y}:\lambda)|^{-1}.$$

An estimate of λ can be obtained as the value maximizing the marginal likelihood function. Structural statements can be made about θ conditional on any λ value.

The case with normal error has been analyzed by Box and Cox (1964) using a likelihood method and a Bayesian method. The likelihood method maximizes the full likelihood function of Y over variation in the regression parameters leaving a *residual likelihood* function for λ:

$$L_1(\lambda \mid \mathbf{y}) = |J(\mathbf{y} \mid \lambda)|/s^n(Y^\lambda).$$

The Bayesian method uses a prior distribution obtained by an approximation and integrates out the regression variables leaving a likelihood factor

$$L_2(\lambda \mid \mathbf{y}) = |J(\mathbf{y} \mid \lambda)|^{(n-r)/n}/s^{n-r}(Y^\lambda).$$

Both of these likelihoods contain contributions from the distribution $g(\theta^{-1}[Y^\lambda]:D^\lambda)\alpha^{-(r+1)}$ for the non informative variable $[Y^\lambda]$.

REFERENCES

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. B* **26** 211–252.

Fraser, D. A. S. (1966). Structural probability and a generalization. *Biometrika* **53** 1–9.