Fibre analysis and tangent models
D. A. S. Fraser

Received: Dec. 17, 1985; revised version: Feb. 24, 1987

Available methods for deriving tests and confidence regions for multi-parameter models seem seriously limited in their feasibility. This paper gives preliminary survey to some developing techniques that use conditioning as a primary device to derive tests and confidence regions. The techniques use fibre analysis which involves the conditional determination of OLS (observed level of significance); the techniques are available initially for models with pivotal quantities and for continuous exponential models. The techniques can then be extended to more general models using the notion of tangent model.

1. Introduction

Multiparameter problems present major difficulties for the usual methods of estimation, tests, and confidence regions.

Of course, the normal linear model is an exception. The rotational symmetry of independent normal errors allows the effective separation of the full model into pieces each involving a single location parameter. The analysis of such component models is relatively straightforward and essentially unequivocal with respect to the basic directions in statistical inference.

However, with departure from the normal linear model, difficulties appear for almost all the standard theoretical methods. In practice various pragmatic and exploratory techniques are then typically invoked.
For example, with the linear model and nonnormal error the usual analysis of variance components are commonly used, and then properties of these determined from the viewpoint of robustness, or by the use of moments, or by bootstrap approximations. In contrast to this, the theoretically well-supported procedures based on the structural or transformation-parameter models are constrained by implementation difficulties for the numerical integration, particularly if the number of parameters exceeds 6 or 7; some computer techniques are currently being developed that may double this bound on dimension.

For other linear models and indeed for more general multiparameter models there appear to be just two basic approaches to obtaining methods of analysis. The first involves a pragmatic search for a variable that estimates the parameter component of interest, followed by a determination of an approximating distribution by moments, or by asymptotic analysis, or by bootstrap sampling. The second, which is perhaps more directly implementable, involves maximum likelihood estimation techniques and the calculation of likelihood drops as successive parameter values are tested. The reference distribution for likelihood drop is commonly $\chi^2/2$ but the support for this seems limited, except in the near pure asymptotic or normal cases.

The Bayesian analysis of multiparameter problems faces essentially the same limitations. Accepting whatever validity or appropriateness applies to a prior still leaves a posterior on the many dimensioned parameter space and the same integration difficulties mentioned earlier for the nonnormal conditional analysis of linear models immediately assert themselves.

In this paper we consider initially the transformation model and the continuous exponential linear model. Fibre analysis allows the determination of observed levels of significance (OLS) and leads to confidence bounds for parameters of interest. The needed computing is manageable on super micros now becoming available. The conical test (Fraser, Massam, 1985) provides a macro test for the full parameter and can be seen as a special limiting case of the fibre analysis.

For more general statistical models it is possible to define a tangent model at a parameter value. The tangent model can be of transformation or of exponential type. Fibre analysis can then be used with such a tangent model thus providing tests and confidence regions for the more general cases.
2. Frame of reference

An appropriate objective for statistical inference is the presentation in an accessible form of all the information concerning a parameter that is contained in a model-data combination (Fraser, 1979; Evans, Fraser, Monette, 1985). Whether or not this can be achieved even for such simple cases as with the location or location-scale models may not be clear; nevertheless, the general objective does remain as a goal or ideal to serve as guide in any search for methods of statistical inference.

The underlying technique for confidence theory involves the use of hypothesis testing. For hypothesis testing in the multiparameter case, optimality of some type typically occurs only with the help of reduction principles and only for relatively simple problems usually involving exponential structure. The central core of theory, does not lead to resolutions for most problems.

At some opposite extreme is the exploratory use of the likelihood function. For example, a value for a component parameter can be assessed by examining the drop in profile likelihood when the component parameter value is specified. While likelihood is well acknowledged as a primary measure for a full parameter value, profile likelihood as intrinsically used for likelihood drop can be very misleading. This can be seen clearly in contexts where a separation of variables reveals inappropriate factors; for example, let $y$ have density $f(y|\theta)$ and $z$ independently have density $\sigma^{-n}(\theta) g((z-\phi)/\sigma(\theta))$ where $g(z)$ has mode at the origin; the joint density is $f(y|\theta) g((z-\phi)/\sigma(\theta))/\sigma^{-n}(\theta)$ and the profile likelihood for $\theta$ is $cf(y|\theta)\sigma^{-n}(\theta)$ and is based on $y$ only, whereas the actual likelihood from $y$ is $cf(y|\theta)$; the factor $\sigma^{-n}(\theta)$ is thus inappropriate and it can seriously affect calculations based on the profile likelihood and in particular can have substantial effects on likelihood drop. This example represents a 'marginalization' effect for profile likelihoods.

Also in the exploratory direction is the use of variables obtained by a pragmatic analysis or by asymptotics, and the consequent determination of an approximating distribution by moments, by asymptotic theory or by simulation sampling, typically with the parameter replaced or centered on its maximum likelihood estimate.

In summary we note that for multiparameter problems there is a wide divergence in methodology. Except for the normal linear model, the primary theoretical methods targeted on optimality are typically unsuccessful; while at some opposite extreme the typically invoked exploratory methods are based on pragmatically determined estimates or likelihood characteristics. In some parallel way, any theoretical solutions are judged
adversely in the presence of any nonuniqueness or failure of optimality while, at the opposite extreme, the exploratory methods exist in some broad context of nonuniqueness.

We suggest that not only is there a large separation between the two approaches, the optimality and the exploratory, but also that there is a large separation in the kinds of assessment used for the two approaches—demanding near absoluteness on one side and just plausibility and appeal on the other.

In this paper we survey some techniques that fall in the intermediate region between the two approaches just discussed. As part of this we acknowledge the double standard in the assessments mentioned above, and emphasize something between the extremes. Specifically, some arbitrariness at initial steps will be acknowledged and yet variables will then be extracted that physically measure the quantities under investigation.

The techniques are developed using conditional methods that go beyond the ordinary use of conditionality as developed from Fisher and Birnbaum. Ordinary conditionality can be interpreted as attempting to separate error and effect. The extended conditionality is used to accomplish a further separation of type of effect and magnitude of effect, and then used in addition to extract a component effect from a full apparent effect.

3. Inference for the full parameter: conical test.

We now overview the conical test procedure for determining the observed level of significance (OLS) for testing \( \theta = \theta_0 \) in a statistical model. The procedure involves the essential use of geometrical properties belonging to or ascribed to the statistical model and is outlined in Fraser and Massam (1985). Repeated use of the procedure gives confidence-region bounds in selected directions on the parameter space.

For this we are particularly interested in the multiparameter case and as a first step we examine the calculation of an OLS for a data value \( y^0 \) in relation to a parameter value \( \theta_0 \), in \( \mathbb{R}^r \).

**Example 3.1** Consider the linear model \( y = X\beta + e \) where the error \( e \) is say nonnormal with density \( f(e) \) and \( X \) is of full column rank. The analysis of the transformation model using conditionality, or the structural model as written above using error variables, leads (Fraser, 1979) to the conditional analysis of \( b(y) \) or \( b = b(e) = b(y) - \beta \) given the vector
\[ d = d(y) = d(e) \] of residuals, where \( b(\cdot) \) is taken for simplicity to be, say, the ordinary least squares function. The error variable \( b = b(e) \) (or pivotal variable \( b = b(y) - \beta \)) has conditional density
\[ g(b) = k^{-1}(d) f(Xb + d) |X'X|^{1/2} \]
given the observed residual vector \( d \).

As largely formulated by Fisher, an observed level of significance in relation to a hypothesis, is a measure of the departure of observed data from what is 'expected' under that hypothesis. The present analysis is based on interpreting this in a literal geometrical sense for the hypothesis that \( \beta = \beta_0 \).

First, for this, we need an initial choice of what is 'expected', say \( b_0 \), for the pivotal variable \( b \). One reasonable choice is the maximum density value \( b_0 \) given by \( g(b_0) = \max g(b) \); an alternative is \( b_0 = E(b) \). The former has substantial computational advantages in the higher dimensional cases of interest; also it corresponds to the MLE \( \hat{\beta} \) given by:
\[ b(y^0) - \hat{\beta} = b_0. \]

Next we consider the departure of the data \( y^0 \) from what is 'expected', and record this as a vector in the error or pivotal space:
\[ B^{obs} = b(y^0) - \beta_0 - b_0 = \hat{\beta}(y^0) - \beta_0. \]
An essential ingredient for the conical test (ibid) is the use of vector properties of the pivot space, which in the present example expresses the linearity involved in the relation between \( b(y) \) and \( \beta \) as coordinates for the column vectors in \( X \); this geometry is taken as an explicit part of the given for the analysis.

A third aspect of the conical test is the acceptance of the direction of apparent effect, and the evaluation conditionally then of the magnitude of effect. The conditional probability (OLS) of as great or greater departure, as just recorded, is then available (Fraser and Massam, 1985) by simple change of variable:
\[ OLS(y^0 re \beta_0) = \frac{\int_1^\infty g(uB^{obs} + b_0)u^{r-1}du}{\int_0^\infty g(uB^{obs} + b_0)u^{r-1}du}. \]
Note that this involves just a one-dimensional integration which is easily implementable on a micro-computer. Also note that this conditional OLS is of course also a marginal OLS. For consider the following: let \( p(y | d) \) be the OLS or \( p \)-value of \( y \) given the direction \( d \), and let \( p(y) = p(y | d) \); as a function on \( \mathbb{R}^n \) \( p(y) \) provides an inverse distance measure from what is expected; the amount of null-hypothesis probability outside the contour \( p(y) = p_0 \) is equal \( p_0 \) conditionally given the direction \( d \) and thus is equal to \( p_0 \) marginally.

The preceding OLS can be used to obtain the \( 1 - \alpha \) confidence bound \( \beta = \hat{\beta}(y^0) + c(v)v \) in any direction \( v (|v| = 1) \) from the MLE \( \hat{\beta}(y^0) \):

\[
\alpha = \frac{\int_0^\infty g(-u v + b_0) u^{r-1} \, du}{\int_0^\infty g(-u v + b_0) u^{r-1} \, du};
\]

for this note that the direction is \(-v\) from the bound \( \beta \) toward \( \hat{\beta}(y^0) \). The resulting confidence region is a marginal confidence region as based on the marginal OLS just described.

The preceding example illustrates the essential ingredients for the conical test: the choice of the 'expected' value for the pivotal or error variable; the use of vector properties of the error space; the separation of direction of effect and magnitude of effect; the latter leading directly to the OLS. The conical test can also be used directly for a component parameter provided the intervening integrations are performed; the next example illustrates this.

**Example 3.2** Consider the linear model \( y = X\beta + \sigma e \) where the error \( e \) is nonnormal say with density \( f(e) \) and \( X \) is full rank. The analysis for the transformation or structural model (Fraser, 1979) leads, for inference concerning \( \beta \) to the conditional analysis of

\[
t(e) = \frac{b(y) - \beta}{s(y)} = \frac{b(e)}{s(e)}
\]

given the vector \( d = d(y) = d(e) \) of standardized residuals. The distribution of the error or pivotal variable \( t = t(e) \) is obtained by one-dimensional (corresponding to \( \sigma \)) integration:

\[
g(t) = k^{-1}(d) \int_0^\infty f(s(Xt + d)) s^{n-1} \, ds \, |X'X|^{1/2}
\]
The 'expected' value for $\beta$ can reasonably be taken to be the maximum pivotal-density estimate (MPDE) $\hat{\beta}(y^0) = b(y^0) - s(y^0)t_0$ where $g(t_0) = \max g(t)$. The departure of the data from what is 'expected' is then given by the vector 

$$T^{obs} = \frac{b(y^0) - \beta_0}{s(y^0)} - t_0 = \frac{\hat{\beta}(y^0) - \beta_0}{s(y^0)}.$$

Again the conical test makes essential use of the geometry of the error space to separate the direction of apparent effect and the magnitude of effect. The OLS, conditional and thus marginal, is then given by a further one-dimensional integration

$$OLS = \frac{\int_1^\infty g(uT^{obs} + t_0)u^{r-1}du}{\int_0^\infty g(uT^{obs} + t_0)u^{r-1}du}.$$

The $1 - \alpha$ confidence bound $\hat{\beta}(y^0) + c(v)s(y^0)v$ in any direction $v$ ($|v| = 1$) from the MPDE $\hat{\beta}$ is given by

$$\alpha = \frac{\int_{c(v)}^\infty g(-u_v + t_0)u^{r-1}du}{\int_0^\infty g(-u_v + t_0)u^{r-1}du};$$

note that $v$ points from data to parameter whereas $T$ points from parameter to data; also note that the confidence region is a marginal confidence region.

The conical test procedure can also be used in a Bayesian context where vector properties are assumed given on the parameter space.

**Example 3.3** Let $p(\theta)$ be the posterior density for $\theta$ on $\mathbb{R}^r$. The maximum probability estimate MPE $\hat{\theta}$ is given by $p(\hat{\theta}) = \max p(\theta)$. A value $\theta_0$ for the parameter can be tested by evaluating the probability of being as far as or farther than $\theta_0$ is from the 'expected' $\hat{\theta}$:

$$\int_1^\infty \frac{p(u(\theta_0 - \hat{\theta}) + \hat{\theta})u^{r-1}du}{\int_0^\infty p(u(\theta_0 - \hat{\theta}) + \hat{\theta})u^{r-1}du}$$

Also the $1 - \alpha$ probability bound $\hat{\theta} + c(v)v$ in any direction $v$ ($|v| = 1$)
from \( \hat{\theta} \) can be determined from
\[
\int_{\exp(\nu)}^{\infty} p(u \nu + \hat{\theta}) u^{r-1} du \\
\int_{0}^{\infty} p(u \nu + \hat{\theta}) u^{r-1} du
\]
This provides an alternative to the usual 'bound' that has a given percentage drop in log-likelihood from the MPDE \( \hat{\theta} \) taken in the direction \( \nu \).

The conical test as described above produces a conditional OLS. It is therefore also a marginal OLS as indicated earlier.

For the *normal* linear model the conical test agrees with the usual chi-square test in the known-variance case, and with the usual F-test in the unknown variance case.

The conical test procedure as just described can be used in any context where
1) The error variable is in one-one correspondence with the parameter.
2) The distribution of the pivotal or error variable is accessible for computation.
3) And to give substance to the notion of a "departure as great", the parameter and parameter spaces have vector space properties.

4. Inference for the full parameter: tangent models

The development of the conical test for a parameter value \( \theta_0 \) takes into consideration both the null \( \theta_0 \) distribution and the full alternative with \( \theta \neq \theta_0 \); the actual construction, however, in the context of data makes direct reference just to those alternatives with primary effects in the data direction itself.

In this section we examine some more general situations in which the given model and the available geometry do not give a clear interpretation for effect in the data direction. The technique is to use only a portion of the full alternative, specifically that indicated by the first derivatives with respect to \( \theta \) at \( \theta_0 \). The details involve working with a tangent model generated by these first derivatives.

The simplest tangent model seems to be the exponential tangent model. Consider the model \( f(x | \theta) \), possibly obtained by some preliminary reduction, and suppose that \( \ln f \) is continuous differentiable with respect to \( \theta \) (say, at \( \theta_0 \); hence a locally fixed carrier). Let
\[ s = s(x) = \frac{\partial}{\partial \theta} \ln f(x | \theta) \bigg|_{\theta = \theta_0}, \quad \delta = \theta - \theta_0, \]
\[ g(x | \theta) = \exp\{\delta s(x) - \phi(\delta)\} f(x | \theta_0) \]
\[ g(s | \theta) = \exp\{\delta s - \phi(\delta)\} \bar{f}(s | \theta_0) \quad (4.1) \]

where \( \exp\{-\phi(\delta)\} \) is the implicitly defined norming constant \( \phi(\delta) \) is the cumulant generating function for the \( \theta_0 \)-distribution and \( \bar{f}(s | \theta) \) is the marginal density for \( s \). Then \( g(x | \theta) \) agrees with the original model to the first derivative at \( \theta_0 \), \( g(s | \theta) \) agrees with the marginal model \( \bar{f}(s | \theta) \) to the first derivative at \( \theta_0 \), and the original score \( s \) at \( \theta_0 \) is now the minimal sufficient statistic in the approximating exponential model. Various inference methods are then available, for example, Efron (1978), Amari (1982), Fraser (1985).

**Example 4.1** Consider the model \( f(x | \theta) \) and a test of \( \theta = \theta_0 \). In particular we assume that \( x \) and \( \theta \) have the same dimension and that \( s \) and \( x \) are in one-one correspondence. For the situation where the full-model conical test is unavailable we derive a test using the \( \theta_0 \) tangent model.

First we note that there are vector space properties for both \( s \) and \( \delta \) in the context of the canonical exponential model. We use these properties to derive the **tangent conical test** using the methods in the preceding section:

\[ OLS(s^0 | \theta_0) = \frac{\int \bar{f}(us^0 | \theta_0) u^{r-1} du}{\int_0^{\infty} \bar{f}(us^0 | \theta_0) u^{r-1} du} \]

where \( s^0 = s(x^0) \) is the observed value of \( s(x) \) and only the null distribution is needed for the calculations. For this with the exponential model (4.1) note that \( s = 0 \) gives the MLE \( \hat{\theta} = \theta_0 \); \( s = s^0 \) gives the MLE \( \hat{\theta}(s^0) \) as solution of \( s^0 = \partial \phi(\hat{\theta} - \theta_0) / \partial \delta \); departure of \( \hat{\theta}(s^0) \) from \( \hat{\theta} = \theta_0 \) is assessed in terms of departure of \( s^0 \) from \( s = 0 \). The shift from the space of \( \hat{\theta} \) to that of \( s \) is to use that particular sample space which records the essential vector space linearity of the exponential model.

We summarise briefly the conditions used for an implementation of the preceding tangent model test: the initial variable has the same dimension as the parameter (perhaps achieved by a pivotal quantity and the use
of the marginal density of that pivotal quantity); the sample space for the derived $s$ is accessible by one-one transformation; a one dimensional numerical integration is performed on the space of $s$. The restricting aspect is the first: the dimensionality of the initial variable; techniques for removing this restriction will be discussed subsequently.

We now discuss a second type of tangent model, a transformation tangent model. The transformation models typically are able to describe a more fundamental geometry than the exponential models; see Fraser (1985). In development of the tangent model we will assume that $t = p(x, \theta)$ is a pivotal quantity with density $g(t)$ on the range of $t$; this can be a conditional or marginal density depending on the original context but as written we assume that $x$ and $t$ are in one-one correspondence. We also assume that the dimension of $t$ is equal to that of $\theta$ (techniques for removing this restriction will be discussed subsequently).

For sampling from a real variable and real parameter, the location tangent model may be found in Fraser (1964). Consider a stochastically-increasing distribution with distribution function $F(x, \theta)$ ($F$ is a decreasing function of $\theta$ for each $x$), and a particular parameter value $\theta_0$. The derivative $F'_\theta(x, \theta_0) = \partial F(x, \theta)/\partial \theta|_{\theta = \theta_0}$ records the rate at which the distribution function value shifts at a point $\theta_0$; this can be used to topologically redefine $x$ so that the shift rate in terms of $x$ is equal to that in terms of $\theta$ (at $\theta_0$) so that $\partial x/\partial \theta$ at $\theta_0$ with $F$ constant is equal to 1 for the new variable. This can be shown (ibid.) to lead to a location model; let

$$t(x) = \theta_0 + \int_{x_0}^{x} -F_{\theta}^{-1}(x, \theta_0)F_x(x, \theta_0) \, dx$$

where $F'_{\theta}(x, \theta) = \partial F(x, \theta)/\partial x$ and $x_0$ is say the median of the $\theta_0$ distribution. Then $t(x)$ has a distribution function $G(t, \theta) = F(x(t), \theta)$ that is one-one equivalent to that for $x$ and yet coincides (ibid) to the first derivative at $\theta_0$ with a location distribution $G(t-(\theta-\theta_0), \theta_0)$; special results are available with such a location model. The location (tangent) model can be used for tests and hence confidence intervals using the theory for translation/transformation models (ibid).

We generalize this procedure in the following example:

**Example 4.2** As indicated above we consider the pivotal variable $t = p(x, \theta)$ with density $g(t)$, and assume that $t$ has the same dimension
r as $\theta$ and that appropriate differentiability conditions hold. Rather than work on the space for $x$ as in the simple tangent model case above, we choose to work on the pivot space and thereby eliminate one Jacobian from the calculations.

A first derivative change at $\theta_0$ corresponds to the matrix $V(x) = \frac{\partial p(x, \theta)}{\partial \theta} |_{\theta = \theta_0}$ with column vectors say $v_1(x), \ldots, v_r(x)$. We express these as functions of the pivot coordinates $t = p(x, \theta_0)$ using the one-one correspondence assumed above: $W(t) = V(x); w_j(t) = v_j(x)$. Now consider a particular direction $a (|a| = 1)$ from $\theta_0$; this produces the pivot space change $v(x) = V(x)a$ or equivalently $w(t) = W(t)a$, which defines a vector field on the pivot space as a function of coordinates $x$ or $t$.

The various vector fields may or may not combine to form the infinitesimal generators for a location or transformation group. However, in the special case $r = 1$ the vector field under mild conditions will typically integrate to give a location group and location model (as on the real line, ibid.).

Let $t_0$ be the maximum density point for $g(t) |W(t)|$. Also let $a$ be the direction from $\theta_0$ that generates the vector field that produces a path, say $C$, from $t_0$ to the observed $t^0 = p(x^0, \theta_0)$ relative to the hypothesis $\theta_0$.

For this path or 'direction' $C$ from $t_0$ to $t^0$ we consider the translation model that coincides with the given model along the path $C$. This gives the observed level of significance:

$$OLS(y^0 \text{ re } \theta_0) = \frac{\int_{u_0}^{\infty} g(t(u)) |W(t(u))| u^{-1} \, du}{\int_{0}^{\infty} g(t(u)) |W(t(u))| u^{-1} \, du}$$

where

$$t(u) = t_0 + \int_{0}^{u} W(t(u)) a \, du$$

is an integral equation giving the path $C$ and $t(u_0) = t_0$. 
Application of the OLS for any direction from a maximum pivotal estimate $\hat{\theta}$ gives a corresponding confidence bound in that direction.

This second form of tangent model may have the extra reinforcements that go with transformation models, but its implementation may well be restricted by difficulties in determining the integrals needed in the definition of $C$.

5. Fibre analysis for component parameters

Consider a pivotal or error variable $t = p(x, \theta)$ with density function $g(t)$ and suppose the dimension $r$ of $t$ is equal to that of $\theta$; this variable and its distribution could have been obtained by various devices including the standard reduction of a transformation, or structural model. Or alternatively consider $t = p(x)$ as the canonical variable of an exponential model; in this case the density $g(t, \theta)$ would depend on $\theta$.

As a preliminary simplification we suppose that the parameterization $\theta = (\theta_1, \ldots, \theta_r)'$ has been arranged with the coordinates in the order corresponding to successive generalizations of an initial model, and such that for testing the reverse order $\theta_r, \ldots, \theta_1$ is used; this is the familiar pattern of the ordinary analysis of variance. For some background details see Fraser (1985).

We consider the problem of testing $\theta_r = \theta_{r_0}$. A subsequent test of $\theta_{r-1}$ would then be in the context of an assumed value for $\theta_r = \theta_{r_0}$. In the transformation model case this would introduce a further conditioning and the resulting problem would be as initially described but with $r$ replaced by $r-1$. And in the exponential case this introduces a marginalization giving again the initial type of problem with $r$ replaced by $r-1$. Thus a test of $\theta_r = \theta_{r_0}$ covers generally the pattern for the iterative testing of the $r$ parameters in the present context. For background discussion of such sequenced testing see Fraser (1985) and Fraser and MacKay (1975).

As in Section 3 we distinguish the two cases, the pivotal case as with translation and transformation models, and the exponential case, and we examine them in that order. For more general cases the notion of tangent model as in Section 4 can be invoked and the results of the special cases used; these more general cases will be examined subsequently together with examples.
For the first case consider tests of the hypothesis $\theta_r = \theta_{r_0}$ using the pivotal variable $t = p(x, \theta)$ with density $g(t)$. The hypothesis with data gives access to characteristics of the pivotal variable $t = p(x, (\theta_1, \ldots, \theta_{r-1}, \theta_{r_0}))$: thus $t \in T^0$ where

$$T^0 = \{p(x, (\theta_1, \ldots, \theta_{r-1}, \theta_{r_0})): \theta_i \text{ free, } i = 1, \ldots, r-1\}$$

The set $T^0$ is $r-1$ dimensional and will typically divide the space of $t$. A two-sided observed level of significance is obtained as

$$\text{OLS} = 2 \min(P(T^+), P(T^-)),$$  \hspace{1cm} (5.2)

where $T^+$ and $T^-$ are the regions on a positive, negative side of the dividing $T^0$ and $P(T) = P(t \in T)$ is calculated using the pivotal density.

We note briefly some set theoretic aspects of the preceding. The parameter space $\Omega = \{\theta\}$ is divided by $\theta_r = \theta_{r_0}$ into $\Omega_+ = \{\theta: \theta_r > \theta_{r_0}\}$, $\Omega_0 = \{\theta: \theta_r = \theta_{r_0}\}$, $\Omega_- = \{\theta: \theta_r < \theta_{r_0}\}$. The one-one mapping $t = p(x^0, \theta)$ maps the preceding sets respectively to the sets $T^+, T^0, T^-$. Certain aspects of the analysis are clearer in the translation/transformation case, in particular the underlying full partition structure on the pivot space is independent of the data value (Fraser, 1979, 1985), provided the succession of hypotheses gives restricted models that inherit the group structure. Accordingly we treat again the Example 3.2.

**Example 5.1.** For the linear model $y = X\beta + \epsilon$ with error density $f(\epsilon)$, let $g(t)$ be the conditional (re d), marginal (re s) density of

$$t(e) = \frac{b(y) - \beta}{s(y)} = \frac{b(\epsilon)}{s(\epsilon)}.$$  \hspace{1cm} (5.3)

The hypothesis $\beta_r = \beta_{r_0}$ gives the value $t_r^0 = (b_r(y^0) - \beta_{r_0})/s(y^0)$ of the $r$th coordinate of the pivotal variable. Note that any hypothesis concerning $\beta_r$ and any data lead to a sample space separation on the basis of the $r$th coordinate $t_r$; this is the partition property mentioned just preceding this example. The observed level of significance is given by

$$\text{OLS}(y^0, \beta_{r_0}) = 2 \min\{P(t_r > t_r^0), P(t_r < t_r^0)\}$$  \hspace{1cm} (5.4)

$$= P( |t_r - t_{r_0}| > |t_r^0 - t_{r_0}| : \text{sgn}(t_r^0 - t_{r_0}))$$
where \( P(t_r > t_{r0}) = \frac{1}{2} \) gives the median \( t_{r0} \) of the hypothesis distribution for \( t_r \).

The preceding OLS is intrinsically the same as that determined by the generalized analysis of variance discussed in Fraser (1985). The difficulties of implementation however rest on the \( r \) dimensional integration needed to obtain the marginal density for \( t_r \). As indicated in Section 2 our interest focuses on obtaining conditional tests for \( \beta_r \) in the presence of the nuisance parameters \( \beta_1, \ldots, \beta_{r-1} \).

As a first proposal consider a conditional test given values for \( \beta_1, \ldots, \beta_{r-1} \). This can be formulated in terms of the conditional density \( g(t_r : t_1, \ldots, t_{r-1}) \) or directly in terms of \( g(t) \):

\[
OLS(y^0, \beta_{r0} : t_1, \ldots, t_{r-1}) = \frac{2 \int_{\tau_0}^\infty g(t_1, \ldots, t_r) dt_r}{\int_{-\infty}^{\tau_0} g(t_1, \ldots, t_r) dt_r}
\]  

(5.5)

where \( t_i^0 = (b_i(y^0) - \beta_i) / s(y^0) \) for \( i = 1, \ldots, r-1 \) and \( \int_{\tau_0}^\infty \) designates \( \int_{-\infty}^{\tau_0} \) which ever gives the smaller integrated value.

As a second proposal we note that a common method of eliminating nuisance parameters is to replace them by their MLE values, or say by their MPDE values as discussed in Example 3.2. This amounts to a parametric model bootstrap where the MLE model density is in fact available for numerical integration.

As a third proposal consider a refinement on the preceding obtained by randomly sampling the \( (t_1, \ldots, t_{r-1}) \) values using some support density \( h(t_1, \ldots, t_{r-1}) \); this would lead to a version of importance sampling. Preferably the choice for \( h(t_1, \ldots, t_{r-1}) \) would be some intuitively or analytically based approximation to the marginal density function \( g_{(r-1)}(t_{(r-1)}) \) for \( t_{(r-1)} = (t_1, \ldots, t_{r-1}) \). Each trial in the simulation would give values for

\[
t_{(r-1)}, \quad OLS(y^0, \beta_{r0} : t_{(r-1)}), \quad w(t_{(r-1)})
\]  

(5.6)

where

\[
w(t_{(r-1)}) = \frac{\int_{\tau_0}^\infty g(t) dt_r}{h(t_{(r-1)})}
\]  

(5.7)
is the weight function of the importance sampling.

Note that each OLS arising from (5.6) is a valid OLS for the hypothesis testing problem. It is of course an OLS evaluated for the sampled values of the pivotal variable or, from a different point of view, for given values of the nuisance parameters. The sampling fluctuations of the OLS and the corresponding weights \( w \) would in themselves give information concerning the stability and sensitivity for the testing procedure.

The long run weighted average from the simulation (5.6) would provide an evaluation of (5.4) except for the possibility of some directional reversals if the median of the conditional distribution changed sides with respect to \( t_r^0 \).

A critical examination of the tests based on (5.5) draws attention to their dependence on the choice of parameterization for the nuisance parameters \( \beta_1, \ldots, \beta_{r-1} \). For the parameterization itself some aspects have been discussed in Fraser (1985) as part of a generalized analysis of variance procedure. Aspects of this can be seen in a familiar way by writing

\[
\beta_1 x_1 + \cdots + \beta_{r-1} x_{r-1} + \beta_r x_r = \beta_1' x_1 + \cdots + \beta_{r-1}' x_{r-1} + \beta_r' (x_r - \sum_{i=1}^{r-1} a_i x_i)
\]

where \( \beta_i' = \beta_i + \beta_r a_i \). Either version can be used to formulate the hypothesis testing problem \( \beta_r = \beta_{r_o} \). Note then that the corresponding new \( t \) variables are given by

\[
t_i' = t_i + a_i t_r \quad i = 1, \ldots, r-1
\]

\[
t_r' = t_r
\]

It follows that the test (5.5) in terms of the nuisance-reparameterization can then be written as

\[
\text{OLS}(y^0, \beta_{r_0}, t_1', \ldots, t_{r-1}') = \frac{2 \int_{-\infty}^{t_r^0} g(t_1 + a_1 t, \ldots, t_{r-1} + a_{r-1} t, t) dt}{\int_{-\infty}^{\infty} g(t_1 + a_1 t, \ldots, t_{r-1} + a_{r-1} t, t) dt}
\]

At issue then is the choice of nuisance reparameterization to increase the accuracy and reliability of the conditional tests based on (5.5)
and (5.6). Some recent attention to conditional inference in the presence of nuisance parameters may be found in Cox and Reid (1985) using the information metric on the parameter space to orthogonalize the nuisance parameters. The equations (5.8) and (5.9) suggest here a sample space procedure to target on a similar objective of separating the inference for the primary parameter.

Ideally one would like a reparameterization that made the conditional test (5.10) independent of or at least insensitive to the condition; achieving this would mean that the conditional OLS would be equal or nearly equal to the preferred marginal OLS. One proposal is to transform the pivotal distribution \( g(t) \) to \( g(t' - \text{at}_r) \) where \( \text{a} = (a_1, \ldots, a_{r-1}, 0) \), to obtain \((0, \ldots, 0, 1)\) as an axis of symmetry of the distribution. An approximation to this can be obtained by evaluating the second derivative matrix \( \partial^2 \ln g(t)/\partial t \partial t' \) at the MPD value \( t_o \) and by then choosing the transformation \( t' = t + \text{at}_r \) to obtain \((0, \ldots, 0, 1)\) as an eigenvector. This reparameterization would correspond to orthogonalization in the multivariate normal error case.

The proposal then is to use this reparameterization to accomplish a separation of \( \beta_r \) and then to calculate a conditional OLS by the methods discussed earlier. The details and properties of this conditional approach will be examined separately.

We now briefly consider the exponential model case. Let \( t = p(x) \) be the canonical variable and \( \theta = (\theta_1, \ldots, \theta_r) \) be the canonical parameter; the model is

\[
g(t, \theta) = \exp\{t'\theta - \phi(\theta)\}h(t).
\]

We assume that this marginal density for \( t \) is available; some techniques for working directly with the density for \( x \) will be discussed elsewhere. We consider the hypothesis \( \theta_r = \theta_{r0} \).

The variable \( t_r \), when examined conditionally given \( t_1, \ldots, t_{r-1} \) depends only on \( \theta_r \). There is a uniqueness property for this: sufficiency arguments in Fraser (1979) show that the one-dimensional contours or one-manifolds given \( t_1, \ldots, t_{r-1} \) are unique contours on the sample space that have a distribution depending only on \( \theta_r \). This property leads to the sequential analysis of variance discussed in Fraser (1985).

The conditional distribution of \( t_r \) is available as
where \( h(t) \) comes from the full density and the normalizing component \( \phi_r(\theta_r) \) is available by one-dimensional integration in effect of the original density \( g(t, \theta) \).

For testing \( \theta_r = \theta_{r0} \) the OLS is given by

\[
\text{OLS}(t_0^0 : t_1^0, \ldots, t_{r-1}^0) = 2 \int_{\text{Est}} t_r^0 \phi_r(t_r, \theta_{r0}) dt_r
\]

\[
= \frac{2 \int_{\text{Est}} t_r^0 \exp\{t_r \theta_{r0}\} h(t_1^0, \ldots, t_{r-1}^0, t_r) dt_r}{\int_{-\infty}^{\infty} \exp\{t_r \theta_{r0}\} h(t_1^0, \ldots, t_{r-1}^0, t_r) dt_r}
\]

A confidence interval for \( \theta_r \) is available by using this OLS and iterating on \( \theta_{r0} \).

As a caution with this procedure we note that a discrete sample space may mean that the conditional sample space can be severely restricted, indeed a single point in an extreme situation.

More positively, we can comment on the uniqueness of the contours for \( \theta_r \) by noting the effects of a nuisance-parameter reparameterization. Let

\[
\theta_i = \theta_0^i + \theta_r a_i \quad \forall \ i = 1, \ldots, r-1;
\]

then

\[
\exp\{\sum_i \theta_i t_i - \phi(\theta)\} h(t)
\]

\[
= \exp\{\sum_i^{r-1} \theta_0^i t_i + \theta_r (t_r + \sum_i^{r-1} a_i t_i) - \phi(\theta)\} h(t)
\]

and we note that the conditioning variable \((t_1, \ldots, t_{r-1})\) remains unchanged while the active variable is subject to a location change which does not affect the OLS defined by (5.13).

6. Some concluding remarks and acknowledgements

In Section 5 we have addressed the statistical problem of testing a primary parameter in the presence of nuisance parameters and examined feasibility for the theoretically indicated tests.

For the pivotal quantity context involving linear parameters the basic distribution is immediately available as a conditional distribution of
dimension equal to the number of parameters. The relevant procedure then for testing a parameter \( \theta_r \) is a one dimensional marginal \( OLS \). Our approach has been to examine a conditional OLS chosen by various means including a type of parameter orthogonalization.

For the exponential model context involving the canonical parameter the basic distribution needed is a marginal distribution of dimension equal to the number of parameters. This marginalization process can inhibit implementation; some conditional procedures will be examined subsequently. The relevant procedure then for testing a parameter \( \theta_r \) is a conditional -distribution OLS that is unique under variable and parameter change.

For the two model types examined the indicated variable for \( \theta_r \) is \( t_r \) and (marginally for the first, conditionally for the second case) involves only \( \theta_r \). The remaining variable \( (t_1, \ldots, t_{r-1}) \) (conditionally for the first case, marginally for the second) involves the nuisance parameters \( \theta_1, \ldots, \theta_{r-1} \) and in general \( \theta_r \). That some information concerning the primary parameter is contained in a likelihood sense in this supplemental variable may be just the price of extracting accessible information.

The author deeply appreciates substantial and helpful suggestions from two referees.

References


Fraser, D.A.S. (1979). Inference and linear Models, Toronto: DAI, University of Toronto Bookroom.

D.A.S. Fraser†
Dept. of Statistics
University of Toronto
Toronto, Ontario M5S 1A1

† Adjunct Professor, York University, University of Waterloo