

# Panel data, local cuts and orthogeodesic models

BENT JESPER CHRISTENSEN<sup>1</sup> and NICHOLAS M. KIEFER<sup>2</sup>

<sup>1</sup>*School of Economics and Management, CAF, CDME, and CLS, University of Aarhus, Bldg 350, DK-8000 Aarhus C, Denmark. E-mail: bjchristensen@econ.au.dk*

<sup>2</sup>*Department of Economics, Cornell University, 490 Uris Hall, Ithaca NY 14853, USA and CAF, CDME, and CLS, University of Aarhus, Bldg. 350, DK-8000 Aarhus C, Denmark*

Orthogeodesic models admit marginal local cuts and therefore separate inference on subparameters is asymptotically justified. Doubly-flat orthogeodesic models admit local cuts marginally and conditionally. Two important empirical models for panel data are used to illustrate this property and demonstrate its usefulness.

*Keywords:* conditional inference; financial economics; labour supply; likelihood; separate inference

## 1. Introduction

Orthogeodesic models were introduced by Barndorff-Nielsen and Blæsild (1993) as a class of statistical models characterized by purely geometric properties. Christensen and Kiefer (1994) introduced local cuts to allow justification of separate inference – conditionally and marginally – in a wide class of statistical models. In this paper we consider two important classes of panel data models and show that they are in the orthogeodesic family and that separate inference on subparameters is justified by the theory of local cuts. This property holds generally: orthogeodesic models admit marginal local cuts.

## 2. Panel data models

In many practical applications of formal statistical models, observations are organized in *panels* with both a time dimension and a cross-sectional dimension. We consider leading cases of economic applications.

**Example 2.1.** *Gaussian panel.* In this example we consider an empirical model from many areas of applied economic analysis. For each individual  $i$ ,  $i = 1, \dots, I$ , we have data  $y_{ij}$  across a number of time periods  $j = 1, \dots, J$ , and we wish to relate these to observed regressors  $z_{ij}$ . Here we focus on a well-known life-cycle labour supply model. The consumer is assumed to maximize a lifetime utility function in the form  $\sum_{t=0}^T \beta^t u(c_t, \ell_t)$ , where  $\beta$  is a discount factor and  $u$  is a utility function increasing in its arguments, consumption  $c$  and leisure  $\ell$ . The maximization is subject to the budget constraint that the value of consumption

over the life cycle is equal to the value of labour income plus the value of initial assets. The resulting consumption and labour supply functions are in each period functions of that period's prices and a time-independent, unobservable variable  $\lambda$  (the marginal utility of wealth) which incorporates the effects of initial assets and prices in all other periods. With suitable further assumptions the model fits in the framework

$$y_{ij} = \alpha_i + \beta z_{ij} + u_{ij}, \tag{2.1}$$

where  $u_{ij}$  is a zero-mean error term. Often, a Gaussian distribution is adopted for the errors. The parameter of interest is the slope  $\beta$ , but unobserved heterogeneity enters through the coefficients  $\alpha_i$ . We are interested in broad panels  $Y_{IJ} = \{y_{ij}, z_{ij} : i = 1, \dots, I, j = 1, \dots, J\}$ , and asymptotics for  $I \rightarrow \infty$ , including in the case where  $J$  is fixed, and in this situation there is no hope of estimating  $\alpha$  precisely.

**Example 2.2 Inverse Gaussian panel.** In this example we consider an empirical model from financial economics. Our key interest is in asymmetry of information, and this provides a link to banking and macroeconomics (Greenwald *et al.* 1984; Gertler 1988). At each time  $t$ , a bank has assets  $A_t$  and liabilities  $L_t$ , and the net worth (equity) is  $N_t = A_t - L_t$ . The assets follow the Itô process

$$dA_t = \eta dt + \zeta dW_t, \tag{2.2}$$

where  $\{W_t\}$  is a standard Wiener process. Since the liabilities represent deposits that are fixed in value we specify  $L_t \equiv L > 0$ , so that net worth evolves according to  $dN_t = dA_t$ , initiated at  $N_0 = A_0 - L > 0$ . The bank becomes insolvent the first time  $t$  when assets  $A_t$  drop below liabilities  $L$ . In the data, banks are separated into different initial size categories, say  $N_{i0}$ ,  $i = 1, \dots, I$ , and quantities are measured in units of initial net worth, e.g.  $n_{it} = N_{it}/N_{i0}$  and  $\sigma_i = \zeta_i/N_{i0}$ . We assume that  $\sigma_i = \beta z_i$ , where  $z_i$  is observed.

In this setting,  $z_i$  is a measure of the degree of information asymmetry between banks and customers. Since banks have different asset portfolio compositions, we have to allow for different returns on equity. As we are looking in particular at banks at risk of failing, we have in mind return on equity  $\eta_i \leq 0$ , and upon normalization we consider the parameter  $\alpha_i = -\eta_i/N_{i0} \geq 0$ . If we have  $J$  banks in size category  $i$ , the  $j$ th bank follows

$$dn_{ijt} = \beta z_i dW_{ijt} - \alpha_i dt, \tag{2.3}$$

with  $n_{ij0} = 1$ ,  $j = 1, \dots, J$ . Panel data on the times to bank failure are  $T_{IJ} = \{t_{ij}, z_{ij} : i = 1, \dots, I, j = 1, \dots, J\}$ , where  $t_{ij}$  denotes the first hitting time for the set  $\{n_{ijt} = 0\}$ .

For fixed  $t > 0$ , we have  $n_{ijt} \sim N(1 - \alpha_i t, \beta^2 z_i^2 t)$ , so the density at the absorbing barrier  $n_{ijt} = 0$  is

$$p(n_{ijt} = 0; \alpha_i, \beta) = \frac{1}{(2\pi t)^{1/2} \beta z_i} e^{-(\alpha_i t - 1)^2 / (2\beta^2 z_i^2 t)}, \tag{2.4}$$

i.e. the well-known Gaussian, whereas the density of  $t_{ij}$  is in the associated inverse Gaussian form

$$p(t_{ii} = t; \alpha_i, \beta) = \frac{1}{(2\pi t^3)^{1/2} \beta z_i} e^{-(\alpha_i t - 1)^2 / (2\beta^2 z_i^2 t)}, \tag{2.5}$$

closely resembling (2.4) (see Cox and Oakes 1984, p. 22) and correspondingly denoted  $N^{-1}(\alpha_i, \beta z_i)$ .

Suppressing terms not depending on parameters, the log-likelihood for the panel data  $T_{IJ}$  is

$$\ell(\alpha, \beta) = -IJ \log \beta + \frac{J}{\beta^2} \sum_i \frac{\alpha_i}{z_i^2} - \frac{J}{2\beta^2} \sum_i \frac{\bar{u}_i}{z_i^2} - \frac{J}{2\beta^2} \sum_i \frac{\alpha_i^2 \bar{t}_i}{z_i^2}, \tag{2.6}$$

where  $\bar{t}_i = \sum_j t_{ij} / J$  and  $\bar{u}_i = \sum_j t_{ij}^{-1} / J$  are the ordinary and inverse harmonic sample means. The score for  $\alpha_i$  is

$$s_i = \frac{J}{\beta^2 z_i^2} - \frac{J \alpha_i \bar{t}_i}{\beta^2 z_i^2}, \tag{2.7}$$

so  $\hat{\alpha}_i = \bar{t}_i^{-1}$ , yielding the log profile likelihood

$$\tilde{\ell}(\beta) = -IJ \log \beta - \frac{J}{2\beta^2} \sum_i \frac{\bar{v}_i}{z_i^2} \tag{2.8}$$

where  $\bar{v}_i = \bar{u}_i - \bar{t}_i^{-1}$  is positive since the inverse operation is convex, and so by Jensen's inequality the average inverse exceeds the inverse average. The profile score is

$$\tilde{s}(\beta) = -\frac{IJ}{\beta} + \frac{J}{\beta^3} \sum_i \frac{\bar{v}_i}{z_i^2} \tag{2.9}$$

and the profile likelihood equation  $\tilde{s}(\beta) = 0$  produces

$$\hat{\beta} = \left( \frac{1}{I} \sum_i \frac{\bar{v}_i}{z_i^2} \right)^{1/2}. \tag{2.10}$$

The problem is that the profile likelihood is not as well behaved as we would like. There is not even any guarantee of consistency of  $\hat{\beta}$  as  $I \rightarrow \infty$ . To see this, it is useful to review some results from Barndorff-Nielsen (1988), who considered the model without regressors, i.e.  $z_i \equiv 1$ . First, the full panel  $\{t_{ij}\}_{ij}$  may be reduced by B-sufficiency to  $\{(\bar{t}_i, \bar{u}_i)\}_i$ , and since  $(\bar{t}_i, \bar{u}_i)$  is in one-to-one correspondence with  $(\bar{t}_i, \bar{v}_i)$ , we may equally consider  $\{(\bar{t}_i, \bar{v}_i)\}_i$ . Furthermore,  $\bar{t}_i$  and  $\bar{v}_i$  are independent,  $J\bar{t}_i \sim N^{-1}(\alpha_i, \beta z_i)$ , and  $J\bar{v}_i \sim \beta^2 z_i^2 \chi_{J-1}^2$ . It follows that  $E(\bar{v}_i / z_i^2) = \beta^2 (1 - 1/J)$ , and so  $\hat{\beta}$  is biased and inconsistent as  $I \rightarrow \infty$  for  $J$  fixed, in particular  $\hat{\beta} \rightarrow \beta(1 - 1/J)^{1/2}$ .

In a given practical situation, if the analysis has been carried to this stage, remedies of the problems arising in situations such as those illustrated in Examples 2.1 and 2.2 obviously exist. However, the key point is that maximum likelihood cannot be pursued without further analysis. Our quest is for an inference principle that allows us to choose the appropriate objective based on conditions *that can be read directly from the likelihood function*.

### 3. Local cuts

In this section we consider separate inference on individual subparameters, with particular reference to proper cuts (Barndorff-Nielsen 1978) and the local generalization due to Christensen and Kiefer (1994).

Let the model function be  $p(x; \theta)$ , and suppose the parameter  $\theta \in \Theta$  may be decomposed as  $\theta = (\phi, \psi) \in \Phi \times \Psi$  such that at each point  $x \in X$  we have

$$p(x; \theta) = p(x; \phi|s)p(s; \psi) \tag{3.1}$$

for a suitable statistic  $s$ . Then  $s$  is S-sufficient for  $\psi$  and S-ancillary for  $\phi$  (Barndorff-Nielsen 1978, p. 50). This restriction is the essential feature of a proper cut. Separate inference on  $\phi$  and  $\psi$  is indicated, based on the pseudo-likelihoods  $p(x; \cdot|s)$  and  $p(s; \cdot)$ , respectively.

Inferential separation is crucial in graphical interaction models (Frydenberg 1990), many normal theory models (Bellhouse 1990), models that possess nuisance parameters (Kalbfleisch and Sprott 1970), partial likelihood situations (Cox 1975), and numerous other cases, and may sometimes be justified based on other sufficiency criteria, including M-sufficiency and G-sufficiency (Barndorff-Nielsen 1978) and L-sufficiency (Barndorff-Nielsen 1988).

To motivate the localization procedure, note that in many practical situations separate inference is appropriate as long as the factorization (3.1) is satisfied to sufficient order of approximation. The exact cut condition is usually unnecessarily stringent in applications. Examples are given below where the new generalization allows natural inferential separation even when the standard condition is violated.

To define a local cut  $s$ , consider shifting  $\phi$  in a neighbourhood that shrinks with sample size. We are interested in conditions such that the asymptotic consequences of the shift for the marginal distribution  $p(s; \theta)$  are less than when instead  $\psi$  is shifted. Similarly, the conditional  $p(x; \theta|s)$  should be less sensitive to  $\psi$ -shifts than to  $\phi$ -shifts. Proportional errors in (3.1) correspond to additive errors in the logarithmic representation given by

$$\log p(x; \phi + n^{-1/2}\varepsilon, \psi) - \log p(x; \phi, \psi) = \tag{3.2}$$

$$[\log p(x; \phi + n^{-1/2}\varepsilon, \psi|s) - \log p(x; \phi, \psi|s)] + [\log p(s; \phi + n^{-1/2}\varepsilon, \psi) - \log p(s; \phi, \psi)],$$

where  $\varepsilon \neq 0$  is a vector conformable with  $\phi$  and  $n$  is the sample size. Using a vector  $\delta \neq 0$  conformable with  $\psi$ , the consequences of shifting this subparameter are defined by symmetry. Suppose the orders  $s_c(\phi)$ ,  $f_c(\psi)$ ,  $s_m(\psi)$  and  $f_m(\phi)$  are such that

$$\log p(x; \phi + n^{-1/2}\varepsilon, \psi|s) - \log p(x; \phi, \psi|s) = O_p(n^{-s_c(\phi)/2}), \tag{3.3a}$$

$$\log p(x; \phi, \psi + n^{-1/2}\delta|s) - \log p(x; \phi, \psi|s) = O_p(n^{-f_c(\psi)/2}), \tag{3.3b}$$

$$\log p(s; \phi, \psi + n^{-1/2}\delta) - \log p(s; \phi, \psi) = O_p(n^{-s_m(\psi)/2}), \tag{3.3c}$$

$$\log p(s; \phi + n^{-1/2}\varepsilon, \psi) - \log p(s; \phi, \psi) = O_p(n^{-f_m(\phi)/2}). \tag{3.3d}$$

Then the  $\varepsilon$ -shift in  $\phi$  is asymptotically of less consequence for  $p(s; \theta)$  than the  $\delta$ -shift in  $\psi$  if

and only if  $f_m(\phi) > s_m(\psi)$ . In the same sense,  $p(x; \theta|s)$  depends less on  $\psi$  than on  $\phi$  if and only if  $f_c(\psi) > s_c(\phi)$ . Thus,  $f$  indicates the ‘fast’ and  $s$  the ‘slow’ orders, and subscripts  $c$  and  $m$  indicate the conditional and marginal models, respectively.

The two terms in square brackets in (3.2) are of order  $s_c(\phi)$  and  $f_m(\phi)$ , respectively, and the total model function dependence on  $\phi$  could be largely through  $p(s; \theta)$  unless  $f_m(\phi) \geq s_c(\phi)$ . Similar considerations on the  $\delta$ -shift in  $\psi$  lead to the requirement  $f_c(\psi) \geq s_m(\psi)$ . In summary,  $s$  is a *local cut* if

$$f_m(\phi) > s_m(\psi), \quad f_c(\psi) > s_c(\phi), \quad f_m(\phi) \geq s_c(\phi), \quad f_c(\psi) \geq s_m(\psi) \quad (3.4)$$

(for details and a characterization in terms of approximately separated Edgeworth expansions, see Christensen and Kiefer 1994). In applications it turns out to be important to have in addition the notion of a *marginal local cut*, relaxing the second strict inequality to a weak one, i.e. a marginal local cut is defined by the requirements

$$f_m(\phi) > s_m(\psi), \quad f_c(\psi) \geq s_c(\phi), \quad f_m(\phi) \geq s_c(\phi), \quad f_c(\psi) \geq s_m(\psi). \quad (3.5)$$

In special cases where either  $\psi$  or  $\phi$  is not present, the concepts of local ancillarity (Cox 1980) or local sufficiency (McCullagh 1984) lead respectively to principles for conditional inference on  $\phi$ , or marginal inference on  $\psi$ , if the relevant differences in asymptotic orders are at least 1 (for a first-order theory) or 2 (for a second-order theory). Similarly, to quantify the nature of the local property in our case, we say that a local cut is of order  $q$  if

$$f_m(\phi) - s_m(\psi) \geq q, \quad f_c(\psi) - s_c(\phi) \geq q \quad (3.6)$$

in (3.4), and if

$$f_m(\phi) - s_m(\psi) \geq q, \quad (3.7)$$

we have a marginal local cut of order  $q$  in (3.5). We are now led to the associated strong inference principle that in a model which admits a local cut of order at least 1 (2 for the second-order theory), separate inference respectively on  $\phi$  and  $\psi$  from the conditional or the marginal distribution is indicated. This notion is now explored further and illustrated.

## 4. Maximum marginal likelihood

In statistical theory, much interest is focused on the maximum likelihood estimator (MLE)  $\hat{\theta} = \arg \max_{\theta} p(x; \theta)$ , where  $p(\cdot; \cdot)$  is the model function. With the introduction of the notion of a marginal local cut, we are led to consider an associated *strong principle* for marginal inference. In particular, suppose that  $\theta = (\phi, \psi)$  and that the parameter of interest is  $\psi$ . If the model admits a marginal local cut  $s$  of order 1 or higher, inference on  $\psi$  in the marginal distribution of  $s$  is indicated, and we consider the maximum marginal likelihood estimator (MMLE)  $\tilde{\psi} = \arg \max_{\psi} p(s; \phi, \psi)$ . In some cases this depends on  $\phi$ , but it is not highly critical which value is used for  $\phi$ , since by the properties of the marginal local cut the dependence asymptotically wears off relatively fast. In other cases,  $\tilde{\psi}$  does not depend on  $\phi$  at all.

**Example 4.1** *Marginal local cut in Gaussian panel.* Consider again Example 2.1 and define

the time series (group) averages  $\bar{y}_i = \sum_j y_{ij}/J$ ,  $i = 1, \dots, I$ , and similarly for  $\bar{z}_i$ . The model (2.1) in deviations from group means is then

$$\tilde{y}_{ij} = \beta \tilde{z}_{ij} + \tilde{u}_{ij}, \tag{4.1}$$

where  $\tilde{y}_{ij} = y_{ij} - \bar{y}_i$ , and similarly for  $\tilde{z}_{ij}$ . It is natural to draw inference on  $\beta$ , the parameter of interest, in this reduced model. Thus, let  $\tilde{y}_i = (\tilde{y}_{i1}, \dots, \tilde{y}_{iJ})$ , and similarly for  $\tilde{z}_i$  and  $\tilde{u}_i$ . If  $u_i$  are independently and identically distributed draws from the  $J$ -dimensional multivariate normal  $N_J(0, \sigma^2 I_J)$  where  $I_J$  is the  $J \times J$  identity matrix and  $\sigma^2 > 0$ , then the marginal distribution for  $s = \{\tilde{y}_i\}_i$  is easily obtained by noting that  $\tilde{y}_i$  are independent across individuals  $i$  and  $\tilde{y}_i \sim N_J(\tilde{z}_i \beta, \sigma^2 M)$ , where  $M = I_J - 1_J 1_J^T/J$ . Here,  $1_J$  is a  $J$ -vector of 1s. Thus, if  $\sigma^2$  does not depend on  $\alpha$ , the distribution of  $s$  does not involve  $\alpha$  either. Of course,  $\bar{y}_i \sim N(\alpha_i + \bar{z}_i \beta, \sigma^2/J)$ , independently across  $i$ . In the notation of Section 3, we would like the interpretation  $\phi = \alpha$ ,  $\psi = (\beta, \sigma^2)$ . Indeed, the conditional distribution  $p(x; \theta|s)$  of the panel  $x = Y_{IJ}$  given  $s$  may be identified with the marginal distribution of  $\bar{y} = \{\bar{y}_i\}_i$  since  $\bar{y}$  and  $s$  are independent. This distribution depends on both  $\phi$  and  $\psi$  so  $s$  is not a proper cut. Nonetheless, it may be proved that under wide conditions  $s$  is a marginal local cut, and marginal inference on  $\psi$  is indicated, thus providing a principled basis for a separate inference procedure common in practical applications; further details on this example may be found in Christensen and Kiefer (1994).

**Example 4.2** *Marginal local cut in inverse Gaussian panel.* Consider again Example 2.2. By analogy with Example 4.1 above, we wish to draw separate inference on  $\psi = \beta$ , and to this end consider  $\phi = \alpha$  as an infinite-dimensional nuisance parameter. Thus, we specify  $s = \bar{v} = \{\bar{v}_i\}_i$ , and we must verify that  $s$  is a marginal local cut. Again, the conditional distribution  $p(x; \theta|s)$  of the panel  $x = T_{IJ}$  may be identified with the marginal distribution of  $\bar{t} = \{\bar{t}_i\}_i$  since  $\bar{t}$  and  $\bar{v}$  are independent. This distribution depends on both  $\phi$  and  $\psi$ , so  $s$  is not a proper cut. We first derive the orders  $f_c(\psi) = f_c(\beta)$  and  $s_c(\phi) = s_c(\alpha)$  from  $p(\bar{t}; \theta|s) = p(\bar{t}; \theta)$ ,  $\theta = (\alpha, \beta)$  (see (3.3)). We have

$$\log p(\bar{t}; \theta) = -I \log \beta + \frac{J}{\beta^2} \sum_i \frac{\alpha_i}{z_i^2} - \frac{J}{2\beta^2} \sum_i \frac{\bar{t}_i^{-1}}{z_i^2} - \frac{J}{2\beta^2} \sum_i \frac{\alpha_i^2 \bar{t}_i}{z_i^2}, \tag{4.2}$$

and to calculate  $f_c(\beta)$  we first note that to appropriate order  $\log(\beta + n^{-1/2}\delta) - \log \beta = \log(1 + n^{-1/2}\delta/\beta) \approx n^{-1/2}\delta/\beta$  and

$$(\beta + n^{-1/2}\delta)^{-2} - \beta^{-2} = [(\beta + n^{-1/2}\delta)\beta]^{-2} [\beta^2 - (\beta + n^{-1/2}\delta)^2] \approx -\beta^{-4} [2n^{-1/2}\delta\beta + n^{-1}\delta^2],$$

so that

$$\begin{aligned} \log p(\bar{t}; \alpha, \beta + n^{-1/2}\delta) - \log p(\bar{t}; \alpha, \beta) \approx \\ -I \frac{n^{-1/2}\delta}{\beta} + J \frac{2n^{-1/2}\delta\beta + n^{-1}\delta^2}{2\beta^4} \left( \sum_i \frac{\bar{t}_i^{-1} - \alpha_i}{z_i^2} + \sum_i \alpha_i \frac{\alpha_i \bar{t}_i - 1}{z_i^2} \right). \end{aligned} \tag{4.3}$$

Since  $\bar{t}_i \sim N^{-1}(\alpha_i, \beta z_i/J^{1/2})$ , it is well known (see, for example, Johnson and Kotz 1970, p. 140) that  $E\bar{t}_i^{-1} = \alpha_i + \beta^2 z_i^2/J$  and  $E\bar{t}_i = 1/\alpha_i$ . Upon combining the first term on the right-

hand side with the first term in parentheses, and ignoring terms of order  $n^{-1}$  and less, (4.3) has been expressed as the sum of  $I$  independent zero-mean random variables. The sample size is  $n = IJ$ , and as  $I \rightarrow \infty$  for  $J$  fixed we obtain easily from a central limit theorem that the total expression is  $O_p(1)$ . It follows from (3.3b) that  $f_c(\beta) = 0$ . Similarly, when perturbing  $\alpha = \{\alpha_i\}_i$  by  $\varepsilon = \{\varepsilon_i\}_i$  we have  $(\alpha_i + n^{-1/2}\varepsilon_i)^2 - \alpha_i^2 = 2\alpha_i n^{-1/2}\varepsilon_i + n^{-1}\varepsilon_i^2$ , so that

$$\log p(\bar{t}; \alpha + n^{-1/2}\varepsilon, \beta) - \log p(\bar{t}; \alpha, \beta) = \frac{Jn^{-1/2}}{\beta^2} \sum_i \frac{\varepsilon_i}{z_i^2} - \frac{J}{2\beta^2} \sum_i \frac{\bar{t}_i(2\alpha_i n^{-1/2}\varepsilon_i + n^{-1}\varepsilon_i^2)}{z_i^2}, \tag{4.4}$$

to order  $n^{-1}$  a sum of  $I$  independent zero-mean terms ( $E\bar{t}_i = 1/\alpha_i$ ), and the CLT yields  $f_c(\alpha) = 0$ .

Turning next to the candidate marginal local cut, we have  $J\bar{v}_i/\beta^2 z_i^2 \sim \Gamma((J-1)/2, 1/2)$ , the gamma distribution, so that

$$\log p(\bar{v}; \theta) = -I(J-1)\log\beta - \frac{J}{2\beta^2} \sum_i \frac{\bar{v}_i}{z_i^2}, \tag{4.5}$$

and obviously  $f_m(\alpha) = \infty$ . Much as before, we obtain

$$\log p(\bar{v}; \alpha, \beta + n^{-1/2}\delta) - \log p(\bar{v}; \alpha, \beta) \approx -I(J-1)\frac{n^{-1/2}\delta}{\beta} + J\frac{2n^{-1/2}\delta\beta + n^{-1}\delta^2}{2\beta^4} \sum_i \frac{\bar{v}_i}{z_i^2}, \tag{4.6}$$

and since  $E\bar{v}_i = (1-1/J)\beta^2 z_i^2$ , we have  $s_m(\beta) = 0$  by the central limit theorem. Clearly, by (3.5),  $s = \bar{v}$  is a marginal local cut, and inference on  $\beta$  in the marginal distribution of  $s$  is indicated.

To obtain the MMLE, differentiation yields the marginal score

$$\tilde{s}_m(\beta) = -\frac{I(J-1)}{\beta} + \frac{J}{\beta^3} \sum_i \frac{\bar{v}_i}{z_i^2}, \tag{4.7}$$

and the marginal likelihood equation  $\tilde{s}_m(\beta) = 0$  produces

$$\tilde{\beta} = \left( \frac{J}{J-1} \frac{1}{I} \sum_i \frac{\bar{v}_i}{z_i^2} \right)^{1/2}. \tag{4.8}$$

Clearly, while the MLE is inconsistent, i.e.  $\hat{\beta} \rightarrow \beta(1-1/J)^{1/2} < \beta$  (see Section 2), the MMLE corrects this deficiency, i.e.  $E(\tilde{\beta}^2) = \beta^2$  and  $\tilde{\beta} \rightarrow \beta$ .

## 5. Orthogeodesic models

An important new class of parametric statistical models, termed the orthogeodesic family, has recently been introduced by Barndorff-Nielsen and Blæsild (1993). Assuming the statistical model  $M = \{p(x, \theta), \theta \in \Theta\}$  is a differentiable manifold, the orthogeodesic property is

geometric and may be characterized in general differential geometric terms. The conditions are essentially that:

- (1)  $M$  is a product manifold,  $M = \Phi \times \Psi$ ;
- (2) the factorization is orthogonal with respect to the Fisher information metric;
- (3) when writing  $M = \{M_\phi : \phi \in \Phi\}$ , the restriction of the metric to  $M_\phi$  does not depend on  $\phi$ ;
- (4) in the sense of Amari (1985),  $M_\phi$  is expected  $\alpha$ -geodesic for some  $\alpha \neq 0$ ;
- (5)  $M_\phi$  is expected 1-flat.

In parametric terms, an orthogeodesic model may be defined by the requirement that there exists a reparametrization  $\theta = (\phi, \psi)$  such that this decomposition of  $\theta$  corresponds to a geometric factorization as just outlined, and in this case  $\theta$  is said to be an ortho-affine parameter. Barndorff-Nielsen and Blæsild (1993) show that with  $\phi$  the location and  $\sigma$  the scale, the Student  $t$  and Cauchy models are orthogeodesic with  $\psi$  in the ortho-affine parameter given by  $\sigma^{-c}$  and  $\log \sigma$ , respectively. Here,  $c = 2(df - 1)/(df + 5)$ , with  $df$  denoting the degrees of freedom. Clearly, many other transformation and exponential models of importance in applications are orthogeodesic.

Introducing generic coordinates  $a, b, c, \dots$  for  $\theta$ , along with  $k, \ell, m$  for  $\phi$  and  $r, s, t$  for  $\psi$ ,  $M$  can be considered a Riemannian manifold with metric the expected information  $i(\theta)$  (elementwise  $i_{ab}$ ; elements of the inverse matrix are  $i^{ab}$ ). The tangent space at  $\theta$  is spanned by  $\partial_a f(\theta)$  for any smooth  $f: \Theta \rightarrow \mathbb{R}$ ; affine connections  $\nabla$  on the associated tangent bundle may be characterized in local coordinates by  $\nabla_{\partial a} \partial_b = \Gamma_{ab}^c \partial_c$  defining the upper Christoffel symbols, or, using the lower Christoffel symbols and the expected information, by

$$\Gamma_{abc} = \Gamma_{ab}^d i_{dc}. \tag{5.1}$$

Of particular interest is the coordinate system defined by the loglikelihood derivatives  $\ell_a, \dots$ . The expected 0-connection (the Riemannian connection) is given by

$$\Gamma_{abc}^{(0)} = \frac{1}{2} \{ \partial_b i_{ac} + \partial_a i_{bc} - \partial_c i_{ab} \} \tag{5.2}$$

and the corresponding  $\alpha$ -connections (see Amari 1985) by

$$\Gamma_{abc}^{(\alpha)} = \Gamma_{abc}^{(0)} - \frac{\alpha}{2} T_{abc} \tag{5.3}$$

with  $T_{abc} = E(\ell_a \ell_b \ell_c)$  the expected skewness. A manifold is  $\alpha$ -flat if there exists a parametrization with  $\Gamma_{abc}^{(\alpha)} = 0$ . Theorems 4.1 and 4.2 in Barndorff-Nielsen and Blæsild (1993) show that conditions (1)–(3) in the characterization of the OGM imply  $\Gamma_{rsk}^{(0)} = 0$ ; (1)–(4) imply  $\Gamma_{rsk}^{(\alpha)} = 0$  for all  $\alpha$  and  $T_{rsk} = 0$ . Write

$$p(\hat{\phi}, \hat{\psi} | \phi, \psi, a) = p(\hat{\phi} | \hat{\psi}, \phi, \psi, a) p(\hat{\psi} | \phi, \psi, a) \tag{5.4}$$

with Edgeworth expansions for each factor

$$p(\hat{\phi} | \hat{\psi}, \phi, \psi, a) \approx N(\phi, i_{\phi\phi}^{-1}(\phi, \psi))(1 + Q_{\phi|\psi}), \tag{5.5}$$



$$p(\hat{\psi}|\phi, \psi, a) \approx N(\psi, i_{\psi\psi}^{-1}(\psi))(1 + Q_\psi) \tag{5.6}$$

where  $a$  is ancillary or approximately ancillary and we have used (2) and (3) in specifying the leading normal term. The adjustment terms are

$$Q_{\phi|\psi} = \frac{1}{6n^{1/2}} \{ \kappa^{abc} h_{abc}(\hat{\theta}, i^{-1}) - \kappa^{rst} h_{rst}(\hat{\psi}, i_{\psi\psi}^{-1}) \}, \tag{5.7}$$

$$Q_\psi = \frac{1}{6n^{1/2}} \kappa^{rst} h_{rst}(\hat{\psi}, i_{\psi\psi}^{-1}). \tag{5.8}$$

The covariant Hermite polynomials  $h_{abc}$  are given by Barndorff-Nielsen and Cox (1989, Section 5.7), who also show that the indicated approximations are valid to order  $n^{-1}$ . We are principally concerned with the coefficients

$$\kappa^{abc} = E(\hat{\theta}^a - \theta^a)(\hat{\theta}^b - \theta^b)(\hat{\theta}^c - \theta^c). \tag{5.9}$$

With this machinery to hand we have the following:

**Theorem 1.** *Orthogeodesic models admit second-order marginal local cuts through  $\hat{\psi}$ .*

**Proof.** Clearly the leading normal term in the marginal distribution of  $\hat{\psi}$  does not depend on  $\phi$ . Turning to the coefficients  $\kappa^{rst}$ , use the linear relationships  $\hat{\psi}^r - \psi^r = i^{rs} \zeta_s + O_p(n^{-1})$  (see Barndorff-Nielsen and Cox 1994) to write

$$\kappa^{rst} = i^{ru} i^{sv} i^{tw} T_{uvw}. \tag{5.10}$$

Condition (5) implies that  $\Gamma_{rst}^{(1)} = 0$ , and hence by (5.3)  $T_{uvw} = \partial_v i_{uw} + \partial_u i_{vw} - \partial_w i_{uv}$ , a function only of  $\psi$ . Since  $\phi$  does not appear in  $h_{rst}(\hat{\psi}, i_{\psi\psi}^{-1}(\psi))$ ,  $\hat{\psi}$  is a second-order marginal ( $n^{-1}$ ) local cut.  $\square$

**Corollary 1.** *Models satisfying only (1)–(3) admit a (first-order) marginal local cut through  $\hat{\psi}$ .*

**Proof.** Examine the leading normal term.  $\square$

Thus, invoking the strong inference principle from Section 4, with  $s = \hat{\psi}$ , separate inference on  $\psi$  in orthogeodesic models is indicated. With additional conditions,  $\hat{\psi}$  becomes a local cut (conditionally as well as marginally). We define the doubly-flat orthogeodesic family as satisfying (1)–(5), together with the following (6)  $M_\psi$  is geodesic; and (7)  $M_\psi$  is 1-flat. Then we have:

**Theorem 2.** *Orthogeodesic models also satisfying (6)  $\Gamma_{k\zeta}^{(\alpha)r} = 0$  (submanifolds  $M_\psi$  are geodesic) and (7)  $\Gamma_{\zeta m}^{(1)k} = 0$  ( $M_\psi$  are 1-flat), admit second-order local cuts through  $\hat{\psi}$ .*

**Proof.** Since  $\hat{\psi}$  is a second-order marginal local cut by Theorem 1, it remains to be shown that the conditional distribution of  $\hat{\phi}$  does not depend on  $\psi$  to order  $n^{-1}$ . From Barndorff-Nielsen and Blæsild (1993, Theorem. 4.1),  $\Gamma_{k\zeta}^{(\alpha)r} = 0$  implies  $i_{k\zeta}(\phi, \psi) = i_{k\zeta}(\phi)$  so the

leading normal term does not depend on  $\psi$ . Turning to the adjustment factor, note that (2) allows elimination of like terms and reduction to

$$Q_{\phi|\psi} = \frac{1}{6n^{1/2}} [\kappa^{k\ell m} h_{k\ell m}(\hat{\phi}, i_{\phi\phi}^{-1}) + 3\kappa^{k\ell r} h_{k\ell r}(\hat{\theta}, i^{-1}) + 3\kappa^{krs} h_{krs}(\hat{\theta}, i^{-1})],$$

with

$$\kappa^{k\ell m} = i^{kn} i^{\ell o} i^{mp} T_{nop}, \kappa^{k\ell r} = i^{kn} i^{\ell o} i^{rt} T_{not}, \kappa^{krs} = i^{kn} i^{rt} i^{su} T_{ntu}.$$

Conditions (1)–(4) imply  $T_{ntu}$  and hence  $\kappa^{krs} = 0$ ; the additional condition (6) implies  $T_{not}$  and hence  $\kappa^{k\ell r} = 0$ . Condition (7) implies  $\Gamma_{k\ell m}^{(1)} = 0$  and hence  $T_{nop} = \partial_o i_{np}(\phi) + \partial_n i_{op}(\phi) - \partial_p i_{no}(\phi)$ , a function only of  $\phi$ . The coefficients  $i^{kn}$ , etc., refer only to the  $i_{\phi\phi}(\phi)$  block of  $i$  and do not involve  $\psi$  either. Hence  $\hat{\psi}$  is a second-order local cut.

**Corollary 2.** *Models satisfying (1)–(3) and (6) admit a (first-order) local cut through  $\hat{\psi}$ .*

**Proof.** Examine the leading normal terms. □

Examining the proof of Theorem 2, we can obtain the further result that for orthogeodesic models (without conditions (6) and (7)) the adjustment term implied by the Edgeworth expansion (5.5) in the conditional distribution is linear in  $\hat{\psi}$ . This is obtained by noting that  $\kappa^{krs} = 0$  and hence  $\hat{\psi}$  appears only in the polynomials  $h_{k\ell r}$ , which contain only linear terms in  $\hat{\psi}$ .

It may be conjectured from the above that the orthogeodesic family generalizes the class of models possessing proper cuts, as do the classes of models that admit local cuts and marginal local cuts. Within the exponential family, this is in fact the case. Barndorff-Nielsen and Blæsild (1983) introduce two subfamilies of the exponential family with  $\theta$ -parallel or  $\tau$ -parallel foliations, both of which are orthogeodesic models, and the  $\theta$ -parallel models coincide with the exponential models permitting proper cuts. Of course,  $\tau$ -parallel models admit second-order marginal local cuts. Finally, the class of doubly-flat orthogeodesic models admits full local cuts (marginal and conditional), providing a useful insight into the geometry of local cuts. The class of doubly-flat orthogeodesic models is a strict generalization of the class of models admitting proper cuts. This can be easily seen by noting that higher-order terms that could be added to the Edgeworth expansions (5.5) and (5.6) involve fourth- and higher-order cross cumulants that are not restricted by our requirements on the second- and third-order cumulants.

**Example 5.1** *Gaussian panel as orthogeodesic model.* Considering the normal distribution as the limit of Student  $t$  distributions as  $df \rightarrow \infty$ ,  $N(\phi, \sigma^2)$  is an orthogeodesic model with  $\psi$  in the ortho-affine parameter given by  $\psi = 1/\sigma^2$ . This is the distribution employed in the Gaussian panel of Examples 2.1 and 4.1, and in many empirical applications to lifetime labour supply. In fact, other orthogeodesic specifications are useful in economic applications, too, including the Student  $t$  with  $df < \infty$  as the distribution of stock returns (which are observed to be fatter-tailed than in the Gaussian case), and panel data models with orthogeodesic model errors are natural tools for their analysis.

**Example 5.2** *Inverse Gaussian panel as orthogeodesic model.* The inverse Gaussian distribution  $N^{-1}(\alpha, \beta)$  possesses a  $\tau$ -parallel foliation and so is an orthogeodesic model. In this case the ortho-affine parameter is  $(\phi, \psi) = (\alpha^{-1}, \beta^{-2})$  and even though the scores  $\ell_\phi$  and  $\ell_\psi$  are not independent,  $\ell_\phi$  is independent of the residual from the quadratic regression of  $\ell_\psi$  on  $\ell_\phi$  to order  $O_p(n^{-1})$  (Barndorff-Nielsen and Blæsild 1992). By Theorem 1, the MLE  $\hat{\psi}$  of  $\psi$  is a second-order marginal local cut, and separate inference on  $\psi$  in the marginal distribution of  $s = \hat{\psi}$  is indicated. By the invariance of maximum likelihood,  $\hat{\psi} = \hat{\beta}^{-2}$ , and we may equally consider marginal inference on  $\beta$  in the distribution of  $\hat{\beta}$ . In the inverse Gaussian panel of Examples 2.2 and 4.2,  $J\bar{v}_i/(\beta^2 z_i^2) \sim \Gamma((J - 1)/2, \frac{1}{2})$ , so the marginal log-likelihood based on  $\hat{\beta}$  is

$$\log p(\hat{\beta}; \beta) = -I(J - 1)\log \beta - \frac{IJ}{2\beta^2} \hat{\beta}^2. \tag{5.11}$$

The resulting marginal score is

$$\tilde{s}_m(\beta) = -\frac{I(J - 1)}{\beta} + \frac{IJ}{\beta^3} \hat{\beta}^2, \tag{5.12}$$

thus producing the MMLE

$$\tilde{\beta} = (1 - 1/J)^{-1/2} \hat{\beta}. \tag{5.13}$$

Thus, the desirable inference procedure from Example 4.2, based on the MMLE from the marginal distribution of  $s = \{\bar{v}_i\}_i$ , again results. This is important since  $\beta$  is the risk-shifting parameter in the asymmetric information banking model of Example 2.2. The procedure may in addition be justified based on modified profile likelihood (Barndorff-Nielsen 1988), but the main point is that it obtains simply by treating the MLE of  $\psi$  in the ortho-affine parametrization as a second-order marginal local cut.

The analysis reveals important relationships between local cuts and orthogeodesic models. In particular, orthogeodesic models always allow separate inference via the theory of local cuts. Of course, not all models admitting local cuts are orthogeodesic. Further, unlike ortho-affine parametrizations, local cuts are invariant to smooth reparametrizations of the form  $(\phi, \psi) \rightarrow (\chi(\phi), \omega(\psi))$ . On the other hand, an orthogeodesic model is characterized by the criterion that an ortho-affine parameter *exists*, while other parametrizations of course may be of interest, too.

## 6. Conclusion

Separate inference on parameters of orthogeodesic models is justified on the basis of the theory of local cuts. Our analysis demonstrates the close connection between geometric and inferential aspects of statistical models. The practical relevance of the results is illustrated in two important empirical models for panel data.

## Acknowledgement

We thank CAE, CAF, CDME, CLS, the Danish Social Science Research Council and the NSF for research support.

## References

- Amari, S.-I. (1985) *Differential-Geometric Methods in Statistics*, Lecture Notes in Statist. 28. Heidelberg: Springer-Verlag.
- Barndorff-Nielsen, O.E. (1978) *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.
- Barndorff-Nielsen, O.E. (1988) *Parametric Statistical Models and Likelihood*, Lecture Notes in Statist. 50. Heidelberg: Springer-Verlag.
- Barndorff-Nielsen, O.E. and Blæsild, P. (1983) Exponential models with affine dual foliations. *Ann. Statist.*, **11**, 770–782.
- Barndorff-Nielsen, O.E. and Blæsild, P. (1992) A type of second order asymptotic independence. *J. Roy. Statist. Soc. Ser. B*, **54**, 897–901.
- Barndorff-Nielsen, O.E. and Blæsild, P. (1993) Orthogeodesic models. *Ann. Statist.*, **21**, 1018–1039.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1989) *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1994) *Inference and Asymptotics*. London: Chapman & Hall.
- Bellhouse, D.R. (1990) On the equivalence of marginal and approximate conditional likelihoods for correlation parameters under a normal model. *Biometrika*, **77**, 743–746.
- Christensen, B.J. and Kiefer, N.M. (1994) Local cuts and separate inference. *Scand. J. Statist.*, **21**, 389–407.
- Cox, D.R. (1975) Partial likelihood. *Biometrika*, **62**, 269–276.
- Cox, D.R. (1980) Local ancillarity. *Biometrika*, **67**, 279–286.
- Cox, D.R. and Oakes, D. (1984) *Analysis of Survival Data*. London: Chapman & Hall.
- Frydenberg, M. (1990) Marginalization and collapsibility in graphical interaction models. *Ann. Statist.*, **18**, 790–805.
- Gertler, M. (1988) Financial structure and aggregate economic activity: an overview. *J. Money, Credit and Banking*, **20**, 559–588.
- Greenwald, B., Stiglitz, J.E. and Weiss, A. (1984) Informational imperfections in the capital market and macroeconomic fluctuations. *Amer. Econom. Rev.*, **74**, 194–199.
- Johnson, N.L. and Kotz, S. (1970) *Continuous Univariate Distributions*. New York: Wiley.
- Kalbfleisch, J.D. and Sprott, D.A. (1970) Applications of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. Ser. B*, **32**, 175–208.
- McCullagh, P. (1984) Local sufficiency. *Biometrika*, **71**, 233–244.

Received August 1997 and revised October 1999