JOURNAL OF Econometrics

ELSEVIER Journal of Econometrics 95 (2000) 391–413

www.elsevier.nl/locate/econbase

# The incidental parameter problem since 1948

Tony Lancaster*

*Department of Economics, Box B, Brown University, Providence, RI 02912, USA*

**Abstract**

This paper was written to mark the 50th anniversary of Neyman and Scott's Econometrica paper defining the incidental parameter problem. It surveys the history both of the paper and of the problem in the statistics and econometrics literature. © 2000 Elsevier Science S.A. All rights reserved.

*JEL classification:* B23; C10

*Keywords:* Incidental parameters; Bayes; Nuisance parameters; Maximum likelihood

## 1. Introduction

It was in the winter of 1948 that Econometrica published an issue[1] containing a single paper. It was by the statistician Jerzy Neyman, and his student Elizabeth (Betty) Scott, (NS), and it was entitled, curiously, *Consistent Estimates Based On Partially Consistent Observations.*

This paper is remarkable for a number of reasons,[2] and not only for its title. It was, I believe, the only occasion on which Neyman, arguably the second most influential statistician of the twentieth century, published in an economics

* Corresponding author.

*E-mail address:* tl@pstc3.pstc.brown.edu (T. Lancaster)

[1] Vol. 16, number 1, pp. 1–32.

[2] For instance it, and therefore Econometrica, contained the first known use of the phrase 'Cramér–Rao inequality', according to David (1995).


PII: S0304-4076(99)00044-5

journal.[3] The paper, which alluded to the work of Ragnar Frisch, contained hints that the models it studied were relevant to econometrics, but its influence on the profession seems to have been close to zero.[4] Yet it has been, and continues to be, cited by social scientists, indeed it has been cited by social scientists in a poisson process of rate 5 per year since 1970. In science, as opposed to social science, it was cited 207 times in 1997 alone and was the second most cited of all Neyman's publications.[5] Its results do not appear in the elementary texts, though they are mentioned, in passing, in some of the advanced ones.[6]

This paper tries to provide a survey of *the incidental parameter problem*, linking the instances of the problem that have been recognised in econometrics with the longstanding discussion of the issue in the statistics literature. The excuse, if one is needed, is that the Neyman and Scott paper, which acted as a catalyst to the statistical argument, was published in an econometrics journal. I shall briefly review the paper and its fate in Sections 2 and 3. Section 4 offers my perception of the current state of play with regard to 'the incidental parameter problem' in econometrics. Section 5 offers a similar reading of the statistics literature. In Section 6, I examine how current statistical approaches work out when applied to problems that have occupied econometricians. Consistently with the thrust of this conference, I shall refer to connections between the theoretical issues, textbook discussions, and econometric practice.

## 2. What did the paper do?

The setting is a sequence of independent random variables whose probability laws involve parameters of two types. The first type appears in the probability law of every random variable; the second type appears in the law of only a finite number, possibly one. Parameters of the first type are 'structural'; parameters of the second type are 'incidental', in Neyman and Scott's terminology.[7] In so far as

[3] The most influential figure, R. A. Fisher, published twice (1923, 1935) in economics journals though both were expository rather than research papers.

[4] Morgan's recent (1990) history of econometrics does not mention it.

[5] According to the 1997 science citation index Neyman received 1404 citations to 44 distinct publications. The most cited paper was the Biometrika paper, Neyman and Pearson (1928), with 212 citations.

[6] One might compare Amemiya (1985, p. 120), 'It is hard to construct examples in which the maximum-likelihood estimator is not consistent… and another estimator is. Neyman and Scott have presented an interesting example of this type…' with Poirier (1995, p. 277), 'There are many examples… in which the MLE is inconsistent: Bahadur (1958), Basu (1955), Kraft and Le Cam (1956), Lehmann (1983, pp. 410–412, 420–412), and Neyman and Scott (1948)'.

[7] 'Structural' was not a good choice of word in view of its subsequent overuse in econometrics. After this section I shall call 'common' the parameters NS called structural.

each observation provides information about the parameters of its probability law, information about the incidental parameters stops accumulating after a finite number have been taken. This is what NS mean by 'inconsistent observations'.

By providing examples, Neyman and Scott established two propositions about the estimation of structural parameters by maximum likelihood (ml). The first of these defined 'the incidental parameter problem' and it is on this I shall focus.[8]

*Proposition. Maximum-likelihood estimates of the structural parameters relating to a partially consistent series of observations need not be consistent.*

To show this consider

*Example 1.* Let $\{x_{ij}\}$ be distributed independently such that

$$p(x_{ij}) = \frac{1}{\sigma\sqrt{2\pi}}\exp\{-(x_{ij}-\alpha_i)^2/2\sigma^2\}, \quad i = 1,\dots,s,\ j = 1,\dots,n_i,\ s \to \infty.$$

The $\{\alpha_i\}$ are incidental, they appear in the law of a fixed number – $n_i$ – of random variables; $\sigma^2$ is structural, it appears in the law of every random variable. It is critical here that we are thinking of the $n_i$ as fixed and $s \to \infty$.

To demonstrate Proposition 1 consider maxmimum-likelihood estimates of $\alpha_i$, $\sigma^2$. For simplicity take $n_i = n =$ constant. Of course $\hat{\alpha}_i = \bar{x}_i$ and then

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{s}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2}{sn} \sim \frac{\sigma^2\chi^2(s[n-1])}{sn}$$

with expectation $\sigma^2(n-1)/n$ for every $s$. So $\hat{\sigma}^2$ cannot possibly be consistent for $\sigma^2$.

## 3. The present status of the paper

### 3.1. In statistics

In the statistics literature Example 1 is the standard illlustration of the failure of ml under random sampling from non-identical parent

[8] The second proposition stated 'Even if the maximum-likelihood estimate of a structural parameter is consistent, if a series of observations is only partially consistent, the maximum-likelihood estimate need not possess the property of asymptotic efficiency.' The example was a heteroskedastic linear model (to use anachronistic terminology) with unequal numbers of observations per person. The mean was the common parameter and the incidental parameters were the $\sigma_i^2$. The ml estimator of the mean is consistent but not efficient.

distributions.[9] Example 1 also plays a central role in the discussion of the general problem of nuisance parameters. For example Basu (1977) writes 'The big question in statistics is: How can we eliminate nuisance parameters from the argument? During the past seven decades an astonishingly large amount of effort and ingenuity has gone into the search for reasonable answers to this question'.

The connection between the elimination of nuisance parameters and the inconsistency of ml estimators is as follows. Maximum likelihood can be done in two stages. In the first the ml estimates of the incidental (nuisance)[10] parameters for given values of the common parameters are computed. These will be functions of the common parameters. In the section stage the nuisance parameters in the likelihood are replaced by these functions and the likelihood maximised with respect to the common parameters. The first stage is what econometricians call concentrating the likelihood and statisticians refer to as constructing the profile likelihood. The first-stage eliminates the nuisance parameters from the likelihood by replacing them by their ml estimates given the remaining parameters. So ml can be interpreted as one (of many)[11] ways of eliminating nuisance parameters, but not, with many nuisance parameters, a very good way because of the unsatisfactory properties of the resulting estimates of the common parameters as evidenced by Example 1.

### 3.2. In econometrics

The role of Example 1 is that of a cautionary tale, invoked to warn applied economists away from ml when working with panel data models with fixed effects. For example, Hausman et al. (1984), when studying the dependence of patent filings on R&D expenditure using a panel of firms wrote, '… we cannot simply estimate separate (firm-specific fixed effects)… because for $T$ held fixed and $N$ large we have the incidental parameter problem and maximum likelihood need not be consistent (see Neyman and Scott…)'

But in recent years the problem has largely disappeared from the radar screen of theoretical econometricians. Newey and McFadden's (1994) survey of *Large sample estimation and hypothesis testing* which allocates 21 pages to the consistency of extremum estimators (including ml) does not mention Neyman and Scott or the incidental parameter problem. Powell's (1994) survey of the *Estimation of semiparametric models* in the latest handbook of econometrics volume

[9] Stuart and Ord (1991, vol. 2, Section 18.31), 'the ML method may become ineffective as in (the) example (1)'.

[10] Roughly speaking 'nuisance' parameters are those which are not of primary interest; 'incidental' parameters are nuisance parameters whose number increases with the sample size.

[11] Basu lists 10 ways of eliminating nuisance parameters.

does cite NS in a one page account of Manski (1987) and of Honoré's (1992) methods for panel data with fixed effects.

The issue was more prominent 10 years earlier. In vol. 2 (1984) of the handbook of econometrics the problem received extensive treatment in Chamberlain's survey[12] of *Panel Data* and again in Aigner, Hsiao, Kapteyn and Wansbeek's survey of *Latent variable models in econometrics*. The monograph of Leamer (1978) also gave a subtle discussion of the incidental parameter problem.[13]

The treatment of Leamer and of Aigner et al. took place in the context of a discussion of errors in variables models. Here, in the so-called functional form of the errors in variables model, the incidental parameters are the 'true' values of the exogenous variables. Neyman and Scott actually provided two more examples or illustrations of the incidental parameter problem. Problems 3 and 4 were both errors in variables problems. In 1948 the profession was still undecided how to treat 'errors' in econometric models, as errors in variables or errors in equations. NS may have believed their paper relevant to econometrics because of its errors in variables examples. In the event, of course, the profession took the errors in equations route and errors in variables problems gradually dropped out of sight.[14] This choice may help to explain the curious status of the paper in econometrics. It won't go away but it is not in the mainstream of the subject.[15] Perhaps the incidental parameter problem is not much discussed in theoretical econometrics because the errors in variables problem is not much discussed.

## 4. The present status of the problem in econometrics

The incidental parameter problem is typically seen to arise (only) with panel data models when allowance is made for agent specific intercepts in a regression model. 'Solutions' are advanced on a case by case basis, typically these involve differencing, or conditioning, or use of instrumental variables.

In this section I shall review a few examples of such panel data models and their incidental parameter problem (if any). In the final section of the paper

[12] Chamberlain's paper (1985) in the Heckman and Singer Econometric Society monograph is an excellent summary of econometric state of the art in the mid-1980s.

[13] 'We consider the inferential puzzles that arise in several simple models whose common feature is an excess of uncertain parameters relative to the number of observations'. Leamer (1978, p. 230).

[14] Errors in variables occupied a whole chapter the first edition of Johnston's (1964) textbook or 10% of the text. In Johnston and DiNardo (1997), the latest edition of the book, the subject takes up less than 1% of the text.

[15] Theil's (1971) Principles of Econometrics, a widely used text in its day, exemplifies the attitude of the profession then and now. The Neyman and Scott paper appears in the bibliography but there is no mention of it, or the incidental parameter problem, in the text.

I shall return to these examples and examine what recent developments in mathematical statistics have to say about them.

In these models the observations are independent across agents given agent specific parameters $\{\alpha_i\}$, common parameters, say, $\lambda$, and covariate values. The joint density of the $T$ observations supplied by each agent is usually specified parametrically in terms of $\alpha_i$,$\lambda$ and the covariates. The 'problem' is the inconsistency, as $N \to \infty$, of ml estimators of the structural parameters, $\lambda$.[16]

The main contribution of econometrics has been to emphasize that some or all of the covariates may be 'chosen' by agent $i$ in the light of his knowledge of $\alpha_i$; knowledge not available to the econometrician. This means that economic theory provides a presumption that $\alpha_i$ and $x_{i1},\dots,x_{iT}$ are dependent in the population. This point plays absolutely no role in the statistics literature.

It is, of course, possible for the econometrician to write his model (likelihood function) marginal on the $\{\alpha_i\}$ rather than conditional on them. This is the so-called random effects approach, whereas a likelihood written conditionally on the $\{\alpha_i\}$ is a fixed effects model. To write the likelihood marginal on the alpha's requires that they be integrated from the probability distribution of the data with respect to some (mixing) distribution, conditional on the covariate sequence. The argument of the preceeding paragraph suggests that this distribution must allow for dependence between alpha and the covariates. This has not prevented the development of a large random effects literature, to which the author has contributed, in which the $\{\alpha_i\}$ are presumed independent of the covariates.[17] We shall not discuss the random effects literature.

Some examples of econometric models in which an incidental parameter problem has been recognised and analysed are as follows.

### *4.1. Linear models: Exogenous covariates*

In this model the observations are independent and normal within agents conditional on $\alpha_i$, $\lambda$ and $x_{i1} \cdots x_{iT}$ and such that

$$\mathscr{E}(y_{it}|\alpha_i,\ \beta,\ \sigma^2,\ \{x_{i1},\dots,x_{iT}\}) = \alpha_i + \beta x_{it},$$

$$\mathscr{V}(y_{it}|\alpha_i,\ \beta,\ \sigma^2,\ \{x_{i1},\dots,x_{iT}\}) = \sigma^2.$$

So here $\lambda = (\beta, \sigma^2)$. For example (Mundlak, 1961) $y$ is farm output, the $x$'s are measured inputs, chosen by the farmer in the light of the unmeasured input

[16] $T$, the length of the panel, corresponds to Neyman and Scott's $n_i$; $N$, the width of the panel (number of agents), corresponds to their $s$.

[17] The problems of the (nonparametric) identification of the mixing distribution and of efficient estimation of the common parameter loom large in this strand of literature. See Lancaster (1990) for a survey of some of these issues and Hahn (1994) for a recent contribution.

$\alpha_i$ which is known to the farmer but an unknown parameter as far as the econometrician is concerned. For this model ml is consistent for $\beta$ but inconsistent for $\sigma^2$. The ml estimator of $\beta$ is least squares and can be computed by first differencing the model or, equivalently, working with data measured from their within agent means. The inconsistency of $\hat{\sigma}^2$ is, as in Example 1, easily patched.[18] This setting is, of course, a more general version of Neyman and Scott's example 1; it's their example with covariates.

### *4.2. Count data models*

Here the observations are independent poisson variates given $\alpha_i$, $\lambda$, $x_{i1}, \dots, x_{iT}$ with

$$\mathscr{E}(y_{it}|\alpha_i, \beta, \{x_{i1}, \dots, x_{iT}\}) = \alpha_i \exp\{\beta x_{it}\},$$

So here the common parameter is $\beta$. For example, $y$ is annual patent filings for firm $i$, the $x$'s are past R&D expenditures, chosen by the firm in the light of its knowledge of the unmeasured input into the inventive process, $\alpha_i$, which for the econometrician is an unknown firm specific constant. For this model ml is consistent, $N \to \infty$, for $\beta$ (Lancaster, 1997). The ml estimator maximizes the multinomial likelihood derived from the distribution of $\{y_{i1} \cdots y_{iT}\}$ conditional on their sum. This is because the likelihood can be reparametrized such that it factors into a term, the multinomial likelihood, involving only $\beta$ and a term depending only on a redefined fixed effect – see below.

### *4.3. Duration models*

The only member of this class whose incidental parameter inconsistency has been studied, to my knowledge, is panel weibull data. For example,[19] let the observations for agent $i$ be independent weibull variates given $\alpha_i$, $\lambda$, $x_{i1}, \dots, x_{iT}$.

$$y_{it}|\alpha_i, \beta, \theta, \{x_{i1}, \dots, x_{iT}\} \sim \text{weibull}(\alpha_i \exp\{\beta x_{it}\}, \theta),$$

$$t = 1, \dots, T, \; i = 1, \dots, N.$$

so that $y_{it}$ has hazard $\theta y^{\theta-1}\alpha_i \exp\{\beta x_{it}\}$ and $\lambda = (\beta, \theta)$. For this model, ml is inconsistent for both $\theta$ and $\beta$, the estimate for $\theta$ converging to a number less than the true value. Chamberlain (1985) provided consistent estimators for

---

[18] In example 1 multiply $\hat{\sigma}^2$ by $n/(n-1)$. The fact that the inconsistency of ml in these models is rather trivial has been unfortunate since it has, I think, obscured the general pervasiveness and difficulty of the incidental parameter problem in econometric models.

[19] This is an excessively simple setup. Real panel durations are often sampled in ways that introduce dependence between successive $y$'s.

a number of parametric duration models with fixed effects. These include the weibull, gamma and lognormal. The method in each case was to maximize the likelihood based on the first differences of the log durations.

### 4.4. Dynamic linear models

Consider, as an example of this class, the model

$$y_{it}|y_{i,t-1}, \dots, y_{i1}, y_{i0}, \alpha_i, \rho, \sigma^2 \sim n(\alpha_i + \rho y_{i,t-1}, \sigma^2);$$

$$t = 1, \dots, T, \; i = 1, \dots, N.$$

Here the observations for each agent are dependent and $\lambda = (\rho, \sigma^2)$. The parameter $\rho$ is not constrained to be less than 1 in modulus, and the likelihood is conditioned on the observed initial values, $\{y_{i0}\}$. This model arises, for example, in cross-country growth models addressing the question of growth rate convergence see, for example, Barro and Xavier Sala-i-Martin, 1994 where the $\alpha_i$ are country specific time invariant factors affecting the level of GDP. For this model ml is inconsistent, $N \to \infty$, for $\rho$, $\sigma^2$. The inconsistency for $\hat{\rho}$ is described in Lancaster (1997). The standard 'solution' (Anderson and Hsiao, 1981 and many subsequent authors) is to write down the first differenced version of the model as

$$y_{it} - y_{i,t-1} = \rho(y_{i,t-1} - y_{i,t-2}) + u_{it} - u_{i,t-1}$$

and estimate $\rho$ using instrumental variables with lagged $y$'s as instruments. Such estimators are consistent. They are also very odd. Instrumental variable methods usually introduce new data, the instrument, and new restrictions specifying something about the correlation of instrumental and previously included variables. This application of iv involves neither new data nor new restrictions. The instruments are lagged values of $y$ and the only restrictions used are already embodied in the likelihood. So instrumental variable methods apparently are being used with this model only because econometricians do not know how to use the likelihood correctly!

A likelihood-based consistent estimator is now available (Lancaster, 1997, and further below).

### 4.5. Binary data models: Exogenous covariates

Let $y$ be binary with

$$\mathscr{E}(y_{it}|\alpha_i, \beta, \{x_{i1} \cdots x_{iT}\}) = G(\alpha_i + \beta x_{it}), \quad t = 1, \dots, T, \quad i = 1, \dots, N$$

with $G(.)$ some known distribution function and the observations are stochastically independent for each agent conditional on $\alpha_i$ and $\beta$. Here $\beta$ is the common parameter. For these models ml is generally inconsistent, $N \to \infty$, for $\beta$. For example, when $G(.)$ is logistic – the logit model –, $T = 2$, $x_{i1} = 0$, $x_{i2} = 1$ then $\hat{\beta} \to 2\beta$. (Andersen, 1970; Chamberlain, 1985; Heckman, 1981). A 'solution' is available when $G(.)$ is logistic in which case $S_i = \sum_{t=1}^{T} y_{it}$ is sufficient for $\alpha_i$ when $\beta$ is known. The distribution of the data $y_{i1} \cdots y_{iT}$ conditional on $S_i$ is, therefore, free of $\alpha_i$. The product of such distributions over the $N$ agents provides a (conditional) likelihood whose mode is consistent, $N \to \infty$, for $\beta$. Other choices of $G(.)$ do not yield a sufficient statistic. Manski (1987) proposes a consistent estimator $\beta$ (up to scale) when $G(.)$ is known only to lie in a rather broad class of distribution functions.

### *4.6. Dynamic binary data models*

When $y$ is binary with

$$\mathscr{E}(y_{it}|\alpha_i, \beta, x_{i1}, \ldots, x_{iT}, y_{i,t-1}, \ldots, y_{i1}, y_{i0}\}) = G(\alpha_i + \rho y_{i,t-1} + \beta x_{it}),$$

$$t = 1, \ldots, T, \quad i = 1, \ldots, N$$

with $G(.)$ some known distribution function we have a dynamic version of the binary data model. Here the observations for each agent are dependent and the common parameter is $\lambda = (\rho, \beta)$. No analysis of the ml estimator is available but monte carlo results of Heckman (1981) show that 'maximum-likelihood estimators exhibit considerable bias' when $G$ is normal. A recent paper by Honoré and Kyriazidou (1997) proposes a consistent estimator of $\beta$, $\rho$.

## 5. The present status of the problem in statistics

Mathematical statistics is in almost complete disarray about the treatment of incidental parameters.[20] Apart from outliers who explicitly deny the problem,[21] there are two main approaches, Bayesian and conditional frequentist.[22]

[20] The symposium in Statistical Science, Reid (1995), demonstrates this.

[21] Such as Lindsey (1996, p. 247), who asserts that Neyman and Scott's example 1 is 'an ill specified model'. 'The preferable approach would be to assume random variability in (the incidental parameters)'. Many statisticians implicitly deny the problem by using random effects modelling. This invariably involves taking the $\alpha_i$ as distributed independently of the covariates, an hypothesis most econometricians would see as unwarranted, except when covariates values are randomly allocated, as in a randomized trial.

[22] A reason for the disarray among statisticians may be that the incidental parameter problem is one in which frequentist and Bayesian answers are sharply different.

### 5.1. Conditional frequentist

The main thrust of frequentist work has been to try to separate the problems of inference about the incidental parameters and inference about the common or structural parameters. Let $\alpha$ denote the incidental parameters and $\lambda$ the structural ones for which consistent inference is desired. Let $S$ be a statistic and suppose the likelihood factors as

$$\ell(y|\alpha, \lambda) = \ell_1(S|\alpha)\ell_2(y|S, \lambda). \tag{1}$$

Then inference about $\lambda$ may be based on the second likelihood which has a fixed and finite-dimensional parametrization. If the parameter space for $\lambda$ does not depend on that for $\alpha$ – variation independence – and standard regularity conditions are satisfied this will provide consistent inference for $\lambda$. In this situation $S$ provides a 'cut' in Barndorff–Nielsen (1978) terminology.

If the likelihood does not factor in the original parametrization we may be able to find a reparametrization from $\alpha$, $\lambda$ to $\alpha^*$, $\lambda$ such that the likelihood does factor. Then the same arguments apply and consistent inference can be based on $\ell_2$. When we can achieve such a factoring the presence of incidental parameters does not create inconsistency in maximum-likelihood estimators of common parameters. When (1) applies, possibly after a reparametrization of the incidental parameters (fixed effects), $\alpha$, $\lambda$ are *likelihood orthogonal*[23] and the incidental parameter problem dissolves. Note that if (1) above applies then

$$\frac{\partial^2 \log \ell}{\partial\alpha\,\partial\lambda} = 0. \tag{2}$$

This implies, but is not implied by, block diagonality of the information matrix.

Next, suppose that $\alpha$, $\beta$ cannot be made likelihood orthogonal but that the likelihood factors as

$$\ell(y|\alpha, \lambda) = \ell_1(S|\alpha, \lambda)\ell_2(y|S, \lambda). \tag{3}$$

In this case the distribution of the data conditional on $S$ does not depend on the incidental parameters but does depend on the common parameters. Again, consistent inference can be made from $\ell_2$. Consistency of the maximizer of $\ell_2$ is not so obvious as in the case of full likelihood separation and needs proof.[24]

---

[23] There are several ways of describing this situation. Basu (1977, p. 364), for example, defines parameters as 'unrelated' if they are independent in the prior and the likelihood satisfies (1).

[24] A simple situation where this factoring occurs is the panel logit with $S$ the the number of 'successes' observed for each person. A subtler example is uncensored duration data where $S$ lists the times of failure but not the identifiers of who failed at each time.

Finally, suppose that the likelihood factors as

$$\ell(y|\alpha, \lambda) = \ell_1(S|\lambda)\ell_2(y|S, \alpha, \lambda). \tag{4}$$

Then inference may be made from the marginal distribution of $S$, which is free of the incidental parameter. Inferences based on the factoring of (3) and (4) are sometimes refered to as *partial likelihood* procedures.

In the above classes of problem it is possible either to separate the likelihood or at least to find a component of the likelihood which is free of the incidental parameter. But in many cases likelihood orthogonality cannot be achieved and a factoring of the likelihood as in (3) or (4) cannot be found. The frequentist literature then proposes approximate separation.

Cox and Reid (1987)'s paper is the influential paper here. They write 'A widely used procedure for inference about a parameter in the presence of nuisance parameters is to replace the nuisance parameters in the likelihood function by their maximum-likelihood estimates and examine the resulting profile[25] likelihood as a function of the parameter of interest. This procedure is known to give inconsistent[26] or inefficient estimates for problems with large numbers of nuisance parameters… . We consider an approach to inference based on the conditional likelihood given maximum-likelihood estimates of the orthogonalized parameters'.

Parameter orthogonality corresponds to block diagonality of the information matrix. If $\alpha_i$ is the fixed effect for agent $i$ and $\lambda$ contains the common parameters write $\alpha_i = \alpha(\alpha_i^*, \lambda)$ where $\alpha_i^*$ is to be chosen to be orthogonal to $\lambda$. Thus, we wish to find a reparametrization of the fixed effects such that

$$\mathscr{E}\,\frac{\partial^2 \log \ell_i}{\partial \alpha_i^* \,\partial \lambda} = 0, \tag{5}$$

where $\ell_i$ is the likelihood contribution of agent $i$. Comparing this to (2) shows that parameter orthogonality achieves on average what likelihood orthogonality achieves identically.

Differentiating the log likelihood with respect to $\alpha^*$ and $\lambda$ and taking expectations we find that block diagonality implies that

$$\frac{\partial \alpha_i}{\partial \lambda_j}\mathscr{E}\left(\frac{\partial^2 \log \ell_i}{\partial \alpha_i^2}\right) + \mathscr{E}\left(\frac{\partial^2 \log \ell_i}{\partial \alpha_i \,\partial \lambda_j}\right) = 0, \quad j = 1, \ldots, K.$$

Solutions of these differential equations determine the dependence of $\alpha_i$ on $\lambda$. The orthogonalized fixed effect, $\alpha_i^*$, can be introduced as the arbitrary constant

[25] Concentrated, in econometric terminology.

[26] This is an implicit reference to Neyman and Scott's example 1. NS is, of course, cited in this paper.

of integration. There is no unique orthogonalized fixed effect since any function of such an effect is also orthogonal to the common parameters.

The proposal of Cox and Reid is to base inference about $\lambda$ on the likelihood conditioned on the maximum likelihood estimate of each $\{\alpha_i^*\}$ taking $\lambda$ as known, $\hat{\alpha}_\lambda^*$.[27] When there exists a sufficient statistic for $\alpha^*$, as in the panel logit case, this leads to the likelihood conditioned on this statistic. But in general the exact conditional distribution is difficult. Cox and Reid therefore proposed an approximate formula for the conditional likelihood given by

$$\ell_M(\lambda) = \ell(\lambda, \hat{\alpha}_\lambda^*)|j_{\alpha^*\alpha^*}(\lambda, \hat{\alpha}_\lambda^*)|^{-1/2}, \tag{6}$$

where

$$j_{\alpha^*\alpha^*}(\lambda, \alpha^*) = -\frac{\partial^2 \log \ell(\lambda, \alpha^*)}{\partial \alpha^* \partial \alpha^*},$$

the $\alpha^*\alpha^*$ component of the negative hessian of the single agent log likelihood. The first term on the rhs of (6) is the concentrated or profile likelihood whose maximum, of course, provides the maximum-likelihood estimate of $\lambda$. The second factor modifies the profile likelihood, hence the $M$ suffix on the left. Subsequent theoretical work (Ferguson et al., 1991) has shown that the first derivative of the log of $\ell_M(\lambda)$, the score function for the objective function (6), has a mean which is $O(1/T)$ as $T \to \infty$ when the observations provided by a single agent are independently and identically distributed. This fact provides the basis of a proof that in this case the ($N \to \infty$) inconsistency of the estimator maximizing (6) is $O(1/T^2)$.[28] By contrast the typical incidental parameter inconsistency of ml is $O(1/T)$, as can be seen from the examples, including NS's Example 1, that have been analyzed.

In the next section we shall gives several examples of parameter orthogonalization and approximate conditional inference which will serve to clarify this procedure.

### 5.2. Bayesian

For a Bayesian the treatment of incidental parameters is clear – integrate them from the likelihood with respect to a prior distribution conditioned on all

[27] A surprising number of listeners to talks on on the subject of this article, having followed the argument to this point, propose *maximizing* the likelihood with respect to $\lambda$ and $\alpha^*$. But by the invariance of ml this leads back to the ml estimator of $\lambda$, rendering the orthogonalization pointless. I hasten to add that no participants in the 1998 Madison conference made this mistake.

[28] It seems to the author probable that this may be proved without requiring independence or identical distributions for the single agent data, that is, for autocorrelated panel data with covariates. If this is so we would have a rather general (approximate) solution to the incidental parameter problem in short panels, at least for scalar incidental parameters.

remaining known or unknown parameters. But the issue is how to choose this prior.

For a subjective Bayesian this is not an issue. The prior should accurately represent 'your' views, and there is little more to be said.[29] Berger et al. (1997)[30] provide a more catholic Bayesian view of the elimination of nuisance parameters. 'The elimination of nuisance parameters[31] is a central but difficult problem in statistical inference. It has been formally addressed only in this century since, in the 19th century Bayes–Laplace school of 'Inverse Probability' the problem was not of particular concern; use of the uniform integrated likelihood was considered obvious.' They go on to argue that the desirable way to go is to integrate out the incidental parameters with respect to an uninformative prior – flat prior random effects. This is straightforward in principle, like all Bayesian calculations, but unfortunately does not, in general, solve the econometric issue of consistency. For example, integrating the panel logit (or probit) likelihood with respect to $\pi(\alpha_i) \propto$ constant, yields a marginal posterior distribution for $\lambda$ whose mode is not consistent for $\lambda$. In the dynamic linear model, integrating out the $\alpha_i$ with respect to a uniform prior yields a marginal posterior density for $\rho$, $\sigma^2$ whose mode is inconsistent for these parameters. So in these two cases a naive construction of a 'uniform integrated likelihood' seems unsatisfactory.

A Bayesian might argue that the demand for a prior on the incidental parameter that yields consistent inference about the common parameter is unwarranted; that it places too much emphasis on consistency and neglects other criteria of reasonable inference. But consider the following scenario. Two econometricians sit in adjacent rooms with the same very large amount of data and the same panel data model. The first has attended an econometrics course through the chapter on instrumental variable estimation, and so has never heard of Bayes' theorem. The second is a Bayesian. Then there are models, mentioned in this survey, where the first econometrician, using an 'IV estimator', will discover the correct common parameter with probability 1, but the second will never get it right even though he is correctly computing his marginal posterior distribution. I suspect that many Bayesian econometricians will feel, as I do, that this is intolerable.

---

[29] 'From a subjective Bayesian point of view the problem (of nuisance or incidental parameters) has a trivial solution: simply integrate the joint posterior with respect to the nusiance parameters and work with the resulting marginal distribution of (the common parameters).' Berger et al. (1997).

[30] Statistical Science, forthcoming 1998.

[31] Note the terminology. 'Nuisance' parameters are parameters not of interest to the investigator. Unlike 'incidental' parameters there is no implication that their number is proportional to the sample size. Barndorff–Nielsen (1978, p. 33), prefers incidental to nuisance, finding the latter 'somewhat emotional'.

So if integrating with respect to a uniform prior is the way to go, that 'uniform' prior must be chosen carefully.[32] But the Bayesian literature is disappointingly limited on the question of choosing priors to achieve consistent inference[33] though there exists a small and mostly frequentist literature on the choice of prior to give good frequentist properties. Reid (1996) provides a useful account of this work.

A possible combination of Bayesian and frequentist approaches is to construct the prior on $\beta$ and the orthogonalized fixed effect. The idea here is that information orthogonality provides a justification for prior independence of $\alpha^*$ and $\lambda$. Unfortunately, since $\alpha^*$ is not unique, this does not dictate which function of $\alpha^*$ to assume uniformly distributed. Sweeting (1987), however, gave the following result. Let $\ell(\lambda, \alpha^*)$ be the single agent likelihood written in terms of the common parameter and the orthogonalized fixed effect. Then the (single agent) marginal posterior density of $\lambda$ is

$$\ell(\lambda) = \ell(\lambda, \hat{\alpha}^*_\lambda)|j_{\alpha^*\alpha^*}(\lambda, \hat{\alpha}^*_\lambda)|^{-1/2}(1 + O_p(1/T)).$$

So a uniform prior on an orthogonalized fixed effect produces, to a close approximation, a marginal posterior density for the parameter of interest that is identical to Cox and Reid's approximate conditional likelihood.[34]

## 6. Orthogonalization, partial likelihood and uniform integration

The absence of a method guaranteed to work in a large class of econometric models means that any paper on this subject must be a catalogue of examples. This paper is no exception, so consider how these approaches work in our econometric examples.

### *6.1. Neyman and Scott*

In the Neyman and Scott (1948) Example 1 full likelihood orthogonality cannot be achieved, but the statistic $\bar{x}_i$ is sufficient for $\alpha_i$ so we can achieve the second type of separation. Conditioning the likelihood on $\{\bar{x}_i\}$ yields

$$\ell(\sigma^2) \propto \frac{1}{\sigma^{s[n-1]}}\exp\{-(1/2\sigma^2)\sum_i \sum_j (x_{ij} - \bar{x}_i)^2\}, \tag{7}$$

[32] As Berger et al., argue.

[33] Diaconis and Freedman (1986) is an exception.

[34] In many models that the author has examined there is a choice of orthogonal fixed effect that makes these objective functions identical.

whose mode, the conditional maximum-likelihood estimator, is $\hat{\sigma}^2 n/(n-1)$. So conditional ml supplies the 'obvious' degrees of freedom correction to the ml estimator.

In this model $\alpha_i$ and $\sigma^2$ are already information orthogonal, by a well known property of the normal distribution, so no reparametrization is required. The ml estimate of $\alpha_i$ is $\bar{x}_i$ and conditioning on the latter leads again to the conditional likelihood.

Finally, integrating out each $\alpha_i$ with respect to a uniform prior leads to (7) as the marginal posterior distribution of $\sigma^2$. So all three incidental parameter solutions lead to the same inferences in this case.

### 6.2. *Linear models*: *Exogenous covariates*

The result from the Neyman and Scott example extends right away to the linear homoscedastic normal model considered earlier as long as we measure the covariates from their within person means.[35] The sum $\sum_t y_{it}$ is sufficient for $\alpha_i$ which is information orthogonal to $\beta$. Conditional likelihood, likelihood conditioned on the ml estimate of an orthogonal fixed effect and the marginal posterior after uniform prior integration of the $\{\alpha_i\}$ all provide the same basis for inference about $\beta$, $\sigma^2$. The resulting ml or posterior mode estimates are the 'within' estimator for $\beta$ and the residual sum squares divided by the correct degrees of freedom for $\sigma^2$.

### 6.3. *Panel counts*

Turning to a simpler example, in the panel poisson count model the likelihood contribution of agent $i$ is

$$\ell_i(\alpha_i, \beta) \propto \exp\{-\alpha_i e^{\sum_s e^{\beta x_{is}}}\}\alpha_i^{\sum_s y_{is}} e^{\beta \sum_s y_{is} x_{is}}.$$

This does not factor. But if we orthogonally reparametrize the fixed effect to

$$\alpha_i^* = \alpha_i \sum_t e^{\beta x_{it}},$$

the likelihood is

$$\ell_i(\alpha_i^*, \beta) \propto e^{-\alpha_i^*}(\alpha_i^*)^{\sum_t y_{it}} \prod_t \left[\frac{e^{\beta x_{it}}}{\sum_s e^{\beta x_{is}}}\right]^{y_{it}}, \tag{8}$$

which is in product form (1). So full likelihood orthogonality is attainable. The second factor is the multinomial distribution of the data conditional on their

[35] This is, of course, already a reparametrization of the fixed effect from $\alpha_i$ to $\alpha_i + \bar{x}_i\beta$.

sum. It follows immediately that ml, ml conditional on the sufficient statistic $\sum_t y_{it}$ and a Bayes posterior after integrating the $\alpha_i^*$ with respect to any proper prior exhibiting independence between $\alpha^*$ and $\beta$ all yield, as $N \to \infty$, the same (consistent) answer, namely the (mode of the) multinomial likelihood. For this model there is no incidental parameter problem.

### *6.4. Duration models*

Consider the earlier example of $T$ independent, uncensored weibull durations observed for each of $N$ agents. The hazard at $y$ is $\alpha_i \theta y^{\theta-1} \exp\{\beta x_{it}\}$. An orthogonal fixed effect is provided by

$$\alpha_i = \exp\{\psi(2) + \theta \alpha_i^*\}, \tag{9}$$

where $\psi(.)$ is the digamma function.[36] Integrating out $\alpha_i^*$ with respect to a uniform prior gives a marginal posterior distribution equal to the likelihood that would be formed from the distribution of the first differences of the logarithms of the data. This is a marginal likelihood procedure, corresponding to the factoring of Eq. (4) above. Its mode is consistent for $\beta$, $\theta$.[37]

### *6.5. Dynamic linear models*

In the dynamic linear model discussed above there is neither likelihood orthogonality nor a suitable partial likelihood. And in contrast to the 'static' linear model, $\alpha_i$ is not information orthogonal to $\rho$, $\sigma^2$. But an orthogonal fixed effect can be found. One such is

$$\alpha_i = y_{0i}(1 - \rho) + \alpha_i^* \mathrm{e}^{-b(\rho)}; \qquad b(\rho) = \frac{1}{T} \sum_{t=1}^{T-1} \frac{T-t}{t} \rho^t.$$

Note that the new fixed effect depends on the intial condition, $y_{0i}$, and on the value of $\rho$.

Cox and Reid would have us condition the likelihood on the ml estimator of $\alpha_i^*$ when the remaining parameters are known: Berger might have us integrate out $\alpha_i^*$ with respect to a uniform prior. As it turns out, both procedures lead to the same (consistent) estimator of $\rho$.[38]

---

[36] Orthogonality requires, as always, that the covariates are measured from the agent specific means.

[37] Chamberlain (1985) noted this result, though not, I think, its derivation via uniform integration of an orthogonalized fixed effect. Cox and Reid (1987) gave the orthogonalization reparametrization (9).

[38] The details are given in the unpublished working paper Lancaster (1997).

A uniform prior on the orthogonalized fixed effect implies a prior on the original fixed effect. This prior often turns out to be rather reasonable. Consider, for example an approximately uniform distribution for $\alpha_i^*$ given $\rho$ and $y_{0i}$ which is centered at zero. This implies a similarly uniform distribution for $\alpha_i$ whose mean depends upon both $\rho$ and the initial condition *except when* $\rho = 1$. This seems a not unreasonable prior view.

### 6.6. Panel logit

In the model with

$$\Pr(y_{it} = 1|\alpha_i, \beta, x_{i1} \dots x_{iT}) = \frac{1}{1 + e^{-\alpha_i - \beta' x_{it}}},$$

$\alpha$ is not information orthogonal to $\beta$. An information orthogonal fixed effect[39] is

$$\alpha_i^* = \sum_{t=1}^{T} \Pr(y_{it} = 1 \,|\, \alpha_i, \beta, x_{i1} \dots x_{iT}). \tag{10}$$

whose maximum-likelihood estimator is

$$\hat{\alpha}_i^* = \sum_{t=1}^{T} y_{it}.$$

This is the sufficient statistic for $\alpha_i$ if $\beta$ is given. Proceeding according to Cox and Reid by conditioning the likelihood on $\hat{\alpha}_i^*$ gives the Andersen (1973) conditional likelihood whose maximum is consistent for $\beta$. On the other hand, integrating out $\alpha_i^*$ with respect to a uniform distribution over its (finite) sample space does not yield the conditional likelihood nor does it yield a posterior distribution for $\beta$ whose mode is consistent. So a uniform distribution for $\alpha_i^*$ does not seem exactly right. On the other hand, a uniform distribution for $\alpha_i^*$ implies a prior for $\alpha_i$ which is an equal weights mixture of logistic densities with an overall mean of $\bar{x}_i\beta$ which seems reasonable. It is not known (to this writer) if there exists a prior for $\alpha_i^*$ which will yield consistent inference for $\beta$.

### 6.7. Panel probit

In the probit model

$$\Pr(y_{it} = 1|\alpha_i, \beta, \{x_{i1} \dots x_{iT}\}) = \Phi(\alpha_i + \beta x_{it}), \qquad t = 1, \dots, T,$$

[39] There exists a simple general formula for the orthogonal fixed in all models where the observations for a single agent are stochastically independent and depend on unknowns *only* through $\mu_{it} = \alpha_i + \beta x_{it}$. The panel poisson, logit, probit are of this single index form. This result and its consequences will be explored in a separate paper.

with $\Phi(.)$ the standard normal distribution function. An orthogonal reparametrization is

$$\alpha_i^* = \sum_{t=1}^{T} \int_{-\infty}^{\alpha_i + \beta x_{it}} h(u)\,\mathrm{d}u \quad \text{where } h(u) = \frac{\phi(u)^2}{\Phi(u)\bar{\Phi}(u)}.$$

As with the logit model a uniform distribution for $\alpha_i^*$ leads to a prior for $\alpha_i$ which is an equal weights mixture of densities proportional to $h(u)$ centred at $x_i\beta$. Nothing is currently known about how recent methods work out with this model though a fairly reasonable conjecture is that the posterior mode of the uniformly integrated likelihood has an inconsistency of $O(1/T^2)$.

### 6.8. Semiparametric models with fixed effects

There are a small number of results on panel data inference with likelihoods specified semiparametrically. The likelihood involves an agent specific fixed effect, $\alpha_i$, a common finite-dimensional parameter $\beta$, and a common unknown function $H(.)$. Identifiability of $\beta$ has been a significant issue in this literature. Two such methods are Manski's (1987) application of the maximum score estimator to a panel binary data context, mentioned earlier in Section 4.5, and Honoré's (1992) method for panel observations on censored and truncated data. Both procedures provide orthogonality conditions free of both the incidental parameter and the unknown function, $H(.)$, killing two birds with one stone. Solving the sample analogs of the orthogonality conditions leads to consistent estimation of $\beta$.[40] Neither method has been explicated as a partial (semiparametric) likelihood procedure so it remains unclear how these procedures relate to the literature I have been surveying.[41]

There is, however, one semiparametric method that lies clearly in the partial likelihood framework. This is a 'within agent' version of Cox's method for survival (duration) data.

#### 6.8.1. Panel censored durations

Let $y_{i1}$, $y_{i2}$ be independent with hazard functions $\alpha_i e^{x_{i1}\beta}\lambda_0(y)$, $\alpha_i e^{x_{i2}\beta}\lambda_0(y)$, respectively, so we have a semiparametric proportional hazards model in which $\lambda_0$ is the unknown function $H(.)$. Let both durations have an (independent) right censoring time of $c$. Consider the rank ordering of $y_{i1}$, $y_{i2}$. If both times are censored this ordering is unknown, otherwise it may be deduced. So we are lead to consider $\Pr(y_{i1} > y_{i2} \mid x_{i1}, x_{i2}, \alpha_i, \beta, \lambda_0(.))$ and given that at least one of the

[40] Up to an arbitrary normalization in the binary data case.

[41] Liang and Zeger (1995) give a statisticians perspective on methods based on unbiased estimating equations or orthogonality conditions and relate them to likelihood-based methods.

pair is less than $c$. This probability is $e^{x_{i1}\beta}/(e^{x_{i1}\beta}+e^{x_{i2}\beta})$ so the likelihood for the rank statistic is of logit form. The estimator which maximizes the product of such terms over agents is consistent for $\beta$ under sensible assumptions about the sequence of fixed effects. The unknown function $\lambda_0(.)$ and the incidental parameters $\alpha_i$ vanish from the partial likelihood because they appear as multiplicative factors in the hazard. The restriction to a common censoring time carries no force here since we can always censor the data at the smallest observed censoring time, if any. The restriction that the $y$'s are independent and censoring is independent appears necessary, however, and this rules out various types of sequential accumulation of data.

This approach is essentially the Cox (1972) partial likelihood argument applied 'within agents'. It originates, I believe, with Ridder and Tunali (1989) who studied survival times of children with the same family (agent), that is, the fixed effect was family specific.[42] Note that the baseline hazard, $\lambda_0$, can be agent specific. Though a frequentist method, a Bayesian version of the argument was given by Kalbfleisch (1978).

## 7. Conclusion

Fifty years seems a long time for a problem to stay unsolved,[43] but then may be that is not so unusual in economics.

More seriously, what are the conclusions that I draw from this survey and especially from recent developments in statistics? The main one is the value of finding an orthogonal reparametrization of the fixed effect. To study the form of such an effect has always been instructive and seems to lead to a deeper understanding of the model. From a frequentist point of view it points the way to a more satisfactory version of maximum-likelihood inference. And from a Bayesian point of view it suggests a way of thinking about the construction of a reasonable prior for the fixed effect. In many models uniform integration or approximate conditional inference using an orthogonal parametrization of the fixed effect seems to lead to either exactly or nearly consistent inference for the parameter of inference in short panels.

A final remark bears on the history of the incidental parameter problem. A recent emphasis in econometric theory has been inference for semiparametric models. Such models involve a fixed and finite-dimensional parameter vector and some unknown function, for example $\lambda_0(.)$ in Section 6.8.2. Now consider the examples of Section 4. As I remarked earlier it would have been possible to

[42] It was also briefly discussed in Lancaster (1990, p. 208).

[43] But compare Fermat... .

extend these models by introducing a population joint distribution of $\alpha$ given $x_1, \ldots, x_T$, say $F(\alpha \mid x)$. One would then integrate the likelihood, in the usual random effects manner, with respect to that distribution. The result would be a likelihood whose argments are the common finite dimensional parameter, $\lambda$, and the (unknown) function $F(\alpha \mid x)$ instead of a likelihood function with arguments $\lambda$ and $\alpha_1, \ldots, \alpha_N$. Likelihoods are not unique and the choice of whether to condition on $F$ or on its realizations is a matter of convenience when the main object is inference about $\lambda$. So from this perspective fixed effects panel data modellers are doing semiparametric econometrics, avoiding inference about $F$ by conditioning the likelihood on its realizations.

I conclude that Neyman and Scott wrote the first semiparametric paper in econometric theory. Or, to put it another way, recent work in semiparametric econometrics continues the Neyman and Scott agenda. The incidental parameter problem is alive and well but in disguise.

## Postscript

But didn't Neyman and Scott have a solution? Yes, two of them, but no general answer. The first was to find a collection of 'orthogonality conditions' depending on the data and on $\lambda$ but not upon the incidental parameters $\alpha$ which converge in probability to zero for every $\lambda$ and $\alpha_1$, $\alpha_2, \ldots$ . On this they remark 'Unfortunately, there are cases where this particular method does not work.' That is, there are models in which there can be no such set of conditions (Neyman and Scott, 1948, p. 19) there is no GMM solution. The second was to find an expression for the expected value of the $\lambda$ score equations (Neyman and Scott, 1948, p. 21). If this expectation does not depend on the incidental parameters then it may be subtracted from the score equations after replacing $\lambda$ by its ml estimate. This amounts to bias correcting the ml scores. The method hinges on the highly restrictive hypothesis that the bias of the ml scores does not depend on the incidental parameters. And there the matter stood.

## For further reading

The following references are also of interest to the reader: Ahn and Schmidt (1995); Aigner et al. ,1984; Anscombe 1964; Arellano and Bover, 1995; Baltagi, 1995; Bickel et al. (1993); Chamberlain (1984); Chamberlain (1992); Cox and Hinkley, 1974; Godambe, 1960; Hsiao, 1986; Jeffreys, 1960; Liang, 1987.

## Acknowledgements

This paper was prepared for the conference on the Principles of Econometrics, Madison, WI, May 2, 3, 1998. Thanks to Chuck Manski and Art Goldberger for close reading of an earlier version. I take full responsibility for the final content. I also recall helpful comments and information provided by Arnold Zellner, Stephen Stigler, David Cox, Andrew Chesher and Jinyong Hahn. My students Tiemen Woutersen and Karsten Hansen provided useful comments.

## References

Ahn, S., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68, 5–27.

Aigner, D., Hsiao, C., Kapteyn, A., Wansbeek, T., 1984. Latent variables. In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics, Vol 3. Elsevier, Amsterdam.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Harvard.

Andersen, E.B., 1970. Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society B 32, 283–301.

Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598–606.

Anscombe, F.J., 1964. Normal likelihoods. Annals of the Institute of Statistical Mathematics 26, 1–19.

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error components models. Journal of Econometrics 68, 29–51.

Baltagi, B.H., 1995. Econometric Analysis of Panel Data. Wiley, New York.

Barndorff-Nielsen, O.E., 1978. Information and Exponential Families in Statistical Theory. Wiley, New York.

Barro, R.J., Xavier Sala-i-Martin, 1994. Economic Growth. McGraw-Hill, New York.

Basu, D., 1977. On the elimination of nuisance parameters. Journal of the American Statistical Association 72 (358), 355–366.

Berger, J.O., Liseo, B., Wolpert, R.L., 1997. Integrated likelihood methods for eliminating nuisance parameters. Working paper 97-01. Institute of Statistics and Decision Sciences, Duke University.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A., 1993. Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins, Baltimore, MD.

Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, Vol. 2. North-Holland, Amsterdam.

Chamberlain, G., 1985. Heterogeneity, omitted variable bias, and duration dependence. In: Heckman, J.J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data. Cambridge University Press, Cambridge.

Chamberlain, G., 1992. Comment: sequential moment restrictions in panel data. Journal of Business and Economic Statistics 10, 20–26.

Cox, D.R., 1972. Regression models and life tables. Journal of the Royal Statistical Society B 34, 187–220.

Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman & Hall, London.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference (with discussion). Journal of the Royal Statistical Society B 49 (1), 1–39.

David, H.A., 1995. First(?) occurence of common terms in mathematical statistics. The American Statistician 49 (2), 121–133.

Diaconis, P., Freedman, D., 1986. On the consistency of Bayes estimates. Annals of Statistics 14, 1–67.

Ferguson, H., Reid N., Cox, D.R., 1991. Estimating equations from modified profile likelihood. In: Godambe, V.P. (Ed.), Estimating Functions. Oxford University Press, Oxford.

Fisher, R.A., 1923. Statistical tests of agreement between observation and hypothesis. Economica 3, 139–147.

Fisher, R.A., 1935. The mathematical distributions used in the common tests of significance. Econometrica 3, 353–365.

Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. Annals of Mathematical Statistics 31, 1208–11.

Hahn, J., 1994. The efficiency bound of the mixed proportional hazard model. Review of Economic Studies 61 (4), 607–629.

Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents – R&D Relationship. Econometrica 52 (4), 909–938.

Heckman, J.J., 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.

Honoré, B.E., 1992. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. Econometrica 60 (3), 533–567.

Honoré, B.E., Kyriazidou, E., 1997. Panel data discrete choice models with lagged dependent variables. Mimeo, Department of Economics, Princeton University.

Hsiao, C., 1986. Analysis of Panel Data. Cambridge University Press, Cambridge.

Jeffreys, H., 1960. Theory of Probability, 3rd Edition, Oxford University Press, Oxford.

Johnston, J., 1964. Econometric Methods, McGraw-Hill, New York.

Johnston, J., DiNardo, J., 1997. Econometric Methods. McGraw-Hill, New York.

Kalbfleisch, J., 1978. Non-parametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society B 40, 214–220.

Lancaster, T., 1990. The econometric analysis of transition data. Cambridge University Press, Cambridge.

Lancaster, T., 1997. Orthogonal parameters and panel data. Brown University Working paper, pp. 97–32.

Leamer, E.E., 1978. Specification Searches. Wiley, New York.

Liang, K.-Y., 1987. Estimating functions and approximate conditional likelihood. Biometrika 74 (4), 695–702.

Liang, K.-Y., Zeger, S.L., 1995. Inference based on estimating functions in the presence of nuisance parameters. Statistical Science 10 (2), 158–172.

Lindsey, J., 1996. Parametric Statistical Inference. Oxford University Press, Oxford.

Manski, C., 1987. Semiparametric analysis of random effects linear models from binary panel data. Econometrica 55, 357–362.

Morgan, M.S., 1990. The history of econometric ideas. Cambridge University Press, Cambridge.

Mundlak, Y., 1961. Empirical production function free of management bias. Journal of Farm Economics 43, 44–56.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. (Chapter 36) In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. Elsevier, Amsterdam.

Neyman, J., Pearson, E.S., 1928. On the use and interpretation of certain test criteria for the purposes of statistical inference. Biometrika 20A, 175; 263.

Neyman, J., Scott, E.L., 1948. Consistent estimation from partially consistent observations. Econometrica 16, 1–32.

Poirier, D.J. (1995). Intermediate Statistics and Econometrics. MIT Press, Cambridge, MA.

Powell, J.L. (1994). Estimation of semiparametric models. Chapter 41. In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. Elsevier, Amsterdam.

Reid, N., 1995. The roles of conditioning in inference. Statistical Science 10 (2), 138–157.

Reid, N., 1996. Likelihood and Bayesian approximation methods. In: Bernardo, J.M., Berger, Jo.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 5. Oxford Science Publications, Oxford.

Ridder, G., Tunali, I., 1989. Analysis of related durations: a semi-parametric approach with an application to a study of child mortality in Malysia. Population Research Center Discussion Paper 90-1, NORC and the University of Chicago.

Stuart, A., Ord, J.K., 1991. Kendall's Advanced Theory of Statistics. Oxford University Press, Oxford.

Sweeting, T., 1987. Discussion of Cox and Reid, 1987.

Theil, H. 1971. Principles of Econometrics. Wiley, New York.