

SEMIPARAMETRIC MODELS: PROGRESS AND PROBLEMS

by

Jon A. Wellner
Department of Statistics
University of Washington
Seattle, Washington

April, 1985

SUMMARY

Semiparametric models, models which incorporate both parametric (finite - dimensional) and nonparametric (infinite - dimensional) components, have received increasing use and attention in statistics in recent years. This paper reviews developments in this very large and rich class of models which spans the middle ground between parametric and nonparametric models. Attention is devoted to a preliminary classification of such models with comments on recent work, to lower bounds for estimation, to two potentially useful methods for construction of efficient estimates, and to open problems.

SOMMAIRE

Les modèles semi-paramétriques (fini - dimensionnelles) aussi bien que non - paramétriques (infini - dimensionnelles), ont reçu en statistique d'attention et d'usage croissant dans les années récentes. Dans ce papier on passe en revue développements dans cette classe grande et riche de modèles qui embrasse l'espace intermédiaire entre les modèles paramétriques et non - paramétriques. On fait attention à une classification préliminaire de tels modèles avec commentaire sur du travail récent, aux limites inférieures pour estimation, à deux méthodes potentiellement utiles pour la construction des calculs efficaces, et aux problèmes ouverts.

1. CLASSES OF SEMIPARAMETRIC MODELS

Little effort has been made to classify or categorize semiparametric models. While such an effort may be premature, it may also help to identify related models and aid in developing methods to apply to new problems. The following scheme should be regarded as provisional and temporary.

The classification of models given here has two fundamental categories: *basic models*, and *derived models*. The basic models consist of exponential family models, group models, and transformation models. The derived models include regression models, convolution models, mixing models, censoring models, and biased sampling models. Although this scheme is both redundant and possibly incomplete, it includes all the semiparametric models with which I am now familiar. The rest of this section elaborates on these categories, and provides examples of the models of the various types with some brief comments on recent work.

Basic Models

The following basic models serve as building blocks in the construction of semiparametric models.

A. Exponential family models. These are familiar parametric models with density (with respect to some measure m)

$$p(x, \theta) = c(\theta) \exp\left(\sum_{i=1}^k Q_i(\theta) T_i(x)\right) h(x)$$

for $\theta \in \Theta \subset R^k$, $x \in X \subset R^d$. While these are themselves completely parametric (finitely dimensional) models, they serve as building blocks for many interesting semiparametric models.

B. Group models. (1). The classical parametric model of this type is obtained as follows: suppose that $Y \cong G \cong P_0$, a fixed distribution on X , and let V denote a group of (one to one) transformations on X parametrized by $\theta \in \Theta \subset R^k$. If $v_\theta \in V$, let $X \equiv v_\theta(Y) \cong P_\theta$ for $\theta \in \Theta$.

Examples:

- (a) Location. $X = R^d$, $v_\theta(x) = x + \theta$ with $\theta \in R^d$, and $P_\theta = P_0(\cdot - \theta)$.
 - (b) Elliptic distributions. $X = R^d$, $v_\theta(x) = \theta^{-1/2}x$ where θ is positive definite and symmetric; $G \equiv P_0$ is spherically symmetric on R^d . Then $P = \{P_\theta: \theta \in \Theta\}$ is the P_0 -family of elliptic distributions.
 - (c) Two-sample models. $X = X_0 \times X_0$, $V = V_0 \times V_0$ where V_0 is a group of matrices on X_0 , $\theta = (\mu, \nu) \in \Theta_0 \times \Theta_0 \equiv \Theta$, $Y = (W, Z)$ with $W, Z \cong P_0$ independent, and $X = (v_\mu(W), v_\nu v_\mu(Z))$.
- (2). By letting the distribution P_0 in (1) range over some large class of probability distributions \mathcal{G} small enough to still allow identification of θ , or at least some important functions of θ , yields a semiparametric model

$$P = \{P_{\theta, C}: \theta \in \Theta, C \in \mathcal{G}\}.$$

Examples:

- (a) If $X = R^1$ in 1(a) above and \mathcal{G} is the family of distributions symmetric about 0, P is the classical symmetric location family.
- (b) If X and Θ are as in 1(b) above and \mathcal{G} is the family of all spherical symmetric distributions, then P is the family of all elliptic distributions; see e.g. Bickel (1982).

Derived Models

The following classes of models are all derived from the basic models given above:

D. Regression models. Given a basic model of one of the three types described above, there is a straightforward recipe for constructing related regression models:

1. Start with an exponential family, group or transformation model $P = \{P_{\theta, G}: \theta \in \Theta, G \in \mathcal{G}\}$ where θ is the finite dimensional Euclidean component of the model and G is the nonparametric or infinite-dimensional component of the basic model.
2. Suppose that $Z \cong H$ on R^d .
3. Given $Z = z$, replace θ (or a component thereof) in the basic model by a semiparametric regression function $r(\gamma, z)$ taking values in Θ where $\gamma \in \Gamma \subset$ some R^k . Different forms for r ranging from parametric to nonparametric regression models, with many interesting intermediate semiparametric forms, are possible. For example:
 - (a) Linear model: $r(\gamma, z) = \gamma z$.
 - (a') Exponential linear model: $r(\gamma, z) = \exp(\gamma z)$.
 - (b) Nonlinear: $r(\gamma, z) = r_0(\gamma, z)$ for a fixed known nonlinear function r_0 .
 - (c) Nonparametric: $r(\gamma, z) = r(z)$, with r smooth.
 - (d) Semiparametric: $r(\gamma, z) = \gamma z_1 + r(z_2)$, where $z = (z_1, z_2)$, and r is smooth.
 - (e) Projection pursuit: $r(\gamma, z) = r(\gamma z)$ where $|\gamma| = 1$ and $r: R^1 \rightarrow R^1$ is smooth.
 - (f) Signal-noise: $r(\gamma z)$ where $r: R^1 \rightarrow R^1$ is periodic with period 1 so that γ is a frequency parameter.

Combining various types of regression functions illustrated by (a) - (f) with the basic regression models A, B or C yields a rich collection of regression models, including parametric, semiparametric, and nonparametric models. Stone (1984) gives an interesting survey and further references. A few selected examples with brief comments concerning recent work follow:

Examples:

- (a) Combining basic model A with the regression model D(a) yields linear exponential family regression models; see e.g. Lehmann (1983) Chapter 3, pages 196 - 207.
- (b) Combining the basic model B1(a) where P_0 is normal with D(a) yields classical parametric normal theory regression models; the extension to B2(a) yields semiparametric linear regression models with arbitrary (symmetric) error distributions.
- (c) The basic model B1(a) (with P_0 a fixed distribution on R^1 ; e.g. normal) combined with the semiparametric regression model D(d) leads to a very interesting class of regression models introduced by Engle, Granger, Rice and Weiss (1983) to study effects of weather on electricity demand, and by Wahba (1984), (1985). This model has one nonparametric component, the smooth regression function r . Generalizations with two nonparametric components by allowing the error distribution to be arbitrary are also of interest. A special case has been studied by Schick (1983), while Stone (1984) discusses a spectrum of related regression models.

(1978) and Oakes (1982), and has been generalized by Gill (1984). Related regression models are discussed by Ridder and Verbakel (1983) and Elbers and Ridder (1983).

- (c) Errors in variables models. Suppose that $X = (Y, Z)$ with

$$\begin{aligned} Y_1 &= Z + \varepsilon_1 \\ Y_2 &= \alpha + \beta Z + \varepsilon_2 \end{aligned}$$

where $Z \cong H$ (non-Gaussian) and $\varepsilon \equiv (\varepsilon_1, \varepsilon_2) \cong N(0, \Sigma)$. The resulting mixture model is an *errors in variables regression* model. Consistent maximum likelihood estimates were obtained by Kiefer and Wolfowitz (1956), but lower bounds for estimation of (α, β) together with asymptotically efficient estimates attaining the bounds were first obtained by Bickel and Ritov (1984).

- (d) If $(Y|Z = z) \cong \text{exponential}(z)$ and $Z \cong H$, then

$$P_H(Y \geq y) = \int_0^{\infty} \exp(-yz) dH(z).$$

Estimation of H via nonparametric maximum likelihood methods in this and more general situations has been considered by Laird (1978) and Jewell (1982). While the estimates are known to be consistent, little is known about the efficiency of the estimates or their rate of convergence.

Other results concerning mixing models and efficient estimation have also been obtained by Lambert and Tierney (1984a,b), and by Has'minskii and Ibragimov (1983).

F. Censoring models. These models are derived from other models of one of the above types as follows: Suppose that $X = P_{\theta, G} \in \mathcal{P}$, and suppose that T is a many - to - one function on the sample space \mathcal{X} of X . Then we can observe only $X^* \equiv T(X) \in P_{\theta, G}^*$.

Examples:

- (a) **Mixing.** The mixing models of E are censoring models with $X^* \equiv T(Y, Z) = Y$.
- (b) **Random right censorship.** In this type of censoring, which has received much use in survival analysis, $X^* \equiv (X_1^*, X_2^*) \equiv T(X_1, X_2) \equiv (X_1 \wedge X_2, 1_{[X_1 \leq X_2]})$. Random right censoring meshes extremely well with Cox's proportional hazards regression model as discussed in D(e). On the other hand, however, this type of censoring can make estimation quite difficult. For example, estimation for the linear regression model D(b) with arbitrary right censoring of the dependent variable has been considered by Miller (1976) and by Buckley and James (1979); see Halpern and Miller (1982). Ritov (1984) has, in spite of the difficulties, computed information lower bounds and produced asymptotically efficient estimators achieving the bounds. Tibshirani (1982) considered a version of this censored regression model with the linear (parametric) regression function replaced by a smooth regression function.
- (c) **Convolution.** Here $X^* \equiv T(X_1, X_2) \equiv X_1 + X_2$ where X_1 and X_2 are independent. The traffic model of Branston (1976) discussed in section 3 is a model which results from this convolution type of censoring combined with a simple mixture model.

G. Biased sampling models. Suppose that $X \in P_{\theta, G} \in \mathcal{P}$, a semiparametric model. Then suppose that $K_i(x)$, $i = 1, \dots, s$ is a collection of known non-negative *biasing kernels* and that λ_i , $i = 1, \dots, s$ is a probability distribution on $\{1, \dots, s\}$. Then the *biased sampling distribution* corresponding to $P_{\theta, G}$,

This truncated regression model has been investigated by Bhattacharya, Chernoff, and Yang (1983). Motivated by a controversy in astronomy concerning Hubble's law, they constructed \sqrt{n} -consistent estimators of the regression parameters θ . Further results for this model have been obtained by Jewell (1984b), who also gives additional examples. Jewell (1984a) has also considered estimation for generalizations of this model with $s \geq 2$ corresponding to stratified sampling on the dependent variable Y .

2. BOUNDS FOR ESTIMATION

Our aim in this section is to briefly survey classical (Cramér - Rao) and modern (Hájek - Le Cam) bounds for estimation in "regular" models. The Hájek - Le Cam approach has led to the development of lower bounds for estimation in nonparametric and semiparametric models. Bounds of this type have been established by Beran (1977), Koshevnik and Levit (1976), Levit (1978), Millar (1979, 1983, 1985) Pfanzagl (1982), and Begun et al. (1983). We give a brief introduction to these bounds for semiparametric models at the end of this section. A thorough treatment will be given in the forthcoming monograph by Bickel, Klaassen, Ritov, and Wellner (1986).

Cramér - Rao Lower Bounds

First consider the case of a "regular" parametric model: suppose that X_1, \dots, X_n are iid $P_\theta \in \mathcal{P} \equiv \{P_\theta: \theta \in \Theta\}$ where $\Theta \subset R^d$ is open, that \mathcal{P} is dominated by a (sigma - finite) measure μ on \mathcal{X} , and let $p(\cdot, \theta) \equiv \frac{dP_\theta}{d\mu}$ for $\theta \in \Theta$. Then the classical *log - likelihood* of an observation X is

$$l(\theta, X) \equiv \log p(X, \theta),$$

the *scores vector* \dot{l} is

$$\dot{l}(\theta, X) \equiv \nabla l(\theta, X) = \frac{1}{p(X, \theta)} \left(\frac{\partial}{\partial \theta_1} p(X, \theta), \dots, \frac{\partial}{\partial \theta_d} p(X, \theta) \right)^T,$$

and the *Fisher information matrix* for θ is

$$I(\theta) = E_\theta[\dot{l}(\theta, X)\dot{l}(\theta, X)^T].$$

Assume that $I(\theta)$ is positive definite so that $I(\theta)^{-1}$ exists.

One form of the classical Cramér - Rao inequality for unbiased estimates $a^T \hat{\theta}_n$ of $a^T \theta$, where a is a fixed vector in R^d , is:

$$n \text{Var}_\theta[a^T \hat{\theta}_n] \geq a^T I(\theta)^{-1} a = \sup_{b \in R^d} \frac{(a^T b)^2}{b^T I(\theta) b}. \quad (1)$$

If we focus on estimation of the first component $\theta_1 \in R^1$ of θ , it follows immediately from (1), the definition of $I(\theta)$, and standard L_2 -projection or regression theory that

$$\begin{aligned} n \text{Var}_\theta[\hat{\theta}_1] &\geq \sup_{b \in R^d} \frac{b_1^2}{b^T I(\theta) b} \equiv I^{11}(\theta) \\ &= \frac{1}{\inf_{c \in R^d, c_1=1} E_\theta[\hat{c}_1 - c_2 \hat{c}_2 - \dots - c_d \hat{c}_d]^2} \\ &= \frac{1}{I_{11}(\theta) - I_{12}(\theta)I_{22}^{-1}(\theta)I_{21}(\theta)} \equiv \frac{1}{I_{11}^*(\theta)} \end{aligned} \quad (2)$$

$$l_n(\theta_n) - l_n(\theta) = \log[p_n(X, \theta_n)/p_n(X, \theta)],$$

then $P \equiv \{P_{n,\theta} : \theta \in \Theta\}$ is *locally asymptotically normal* (LAN) at θ if there is a vector of $L_2(P_\theta)$ functions $\underline{l}_n(\theta)$ and a nonsingular matrix $I(\theta)$ such that, with

$$l_n(\theta_n) - l_n(\theta) = \underline{\dot{l}}_n(\theta)^T h - \frac{1}{2} h^T I(\theta) h + R_n(\theta, h), \quad (8)$$

it follows that, in $P_{n,\theta}$ -probability,

- (i) $R_n(\theta, h) \rightarrow_p 0$ uniformly on bounded h -sets, and
- (ii) $\underline{\dot{l}}_n(\theta) \rightarrow_d N(0, I(\theta))$.

Thus $l_n(\theta_n) - l_n(\theta) \rightarrow_d N(-\frac{1}{2}\sigma^2, \sigma^2)$ with $\sigma^2 = h^T I(\theta) h$. In "regular families" P (with iid observations) $\underline{\dot{l}}_n(\theta) = n^{-1/2} \sum_{i=1}^n \dot{\underline{l}}(\theta, X_i)$ where $\dot{\underline{l}}$ is the scores vector (for $n=1$) and $I(\theta)$ is the information matrix.

Because of our interest here in the parametric component θ of a semiparametric model $P = \{P_{\theta,C}\}$, we formulate versions of the convolution and asymptotic minimax bounds for the first component θ_1 of θ .

A sequence of estimators T_{1n} of θ_1 is *regular* at θ if, under P_{θ_n}

$$\sqrt{n}(T_{1n} - \theta_{1n}) \rightarrow_d T_1$$

for every $\theta_n = \theta + n^{-1/2}h$ where the distribution $L(T_1)$ of T_1 does not depend on h .

Theorem 1. (Hájek, 1970). Suppose that P is LAN at θ and that T_{1n} is a regular estimator with limit distribution $L(T_1)$. Then

$$T_1 \cong Z_1 + W_1 \quad (9)$$

where $Z_1 \cong N(0, 1/I_{11}^*(\theta))$, $I_{11}^*(\theta)$ is as in (3), and W_1 is independent of Z_1 .

Thus any regular estimator T_{1n} of θ_1 must have a limit distribution which is at least as dispersed as $N(0, 1/I_{11}^*(\theta))$, and it makes sense to call a regular estimator T_{1n} asymptotically efficient if it converges in distribution to Z_1 ; i.e. if $W_1 \equiv 0$ in (9).

Now suppose that $w: R^1 \rightarrow R^+$ satisfies:

- (i) $w(x) = w(-x)$ for all $x \in R^1$,
- (ii) $w(0) = 0$, $w(x)$ increases in $x \geq 0$,
- (iii) $E w(\sigma Z) < \infty$ for all $\sigma > 0$ where $Z \cong N(0, 1)$.

Theorem 2. (Hájek, 1972). Suppose that P is LAN at θ and that w satisfies (i) - (iii). Then, for any estimator T_{1n} of θ_1 ,

$$\liminf_{M \rightarrow \infty} \liminf_{n \rightarrow \infty} \sup_{\sqrt{n}|\theta_n - \theta| \leq M} E_{\theta_n} w(\sqrt{n}(T_{1n} - \theta_{1n})) \geq E w(Z_1) \quad (10)$$

where $Z_1 \cong N(0, 1/I_{11}^*(\theta))$ as in theorem 1.

If the uniformity in h in (i) of the definition of a LAN family is relaxed to just pointwise convergence, then theorems 1 and 2 continue to hold, but the bounds may not be attainable. Furthermore, if attention is restricted to regular estimates, then (10) holds without the supremum on the left side.

Bounds for Semiparametric Models

The Hájek - Le Cam convolution and asymptotic minimax bounds stated above for a parametric model P_0 continue to hold in a wide range of regular

Method 1: Efficient Score Equation

Suppose that it is possible to calculate the *efficient score function* \mathbf{i}_1^* for θ_1 ,

$$\mathbf{i}_1^* = \mathbf{i}_1 - I_{12}I_{22}^{-1}\mathbf{i}_2 = \mathbf{i}_1 - \Pi(\mathbf{i}_1|\dot{P}_{\theta_2})$$

and the *effective information*

$$I_{11}^*(\theta) = E_{\theta}(\mathbf{i}_1^{*2}).$$

Furthermore, suppose that $\bar{\theta}_n$ is a \sqrt{n} -consistent estimator of θ , $\sqrt{n}(\bar{\theta}_n - \theta) = O_p(1)$. Then define $\hat{\theta}_{1n}$ to be either a solution of the *efficient score equation*

$$\sum_{i=1}^n \mathbf{i}_1^*(\hat{\theta}_{1n}, \bar{\theta}_{2n}, X_i) = 0,$$

or a one-step approximation thereof:

$$\begin{aligned} \hat{\theta}_{1n} &= \bar{\theta}_{1n} + \frac{\frac{1}{n} \sum_{i=1}^n \mathbf{i}_1^*(\bar{\theta}_n, X_i)}{I_{11}^*(\bar{\theta}_n)} \\ &= \bar{\theta}_{1n} + \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{i}}_1(\bar{\theta}_n, X_i) \end{aligned} \quad (1)$$

where $\tilde{\mathbf{i}}_1$ is the efficient influence curve for θ_1 . Additional smoothing may also be required in forming the sums in (1), but we have omitted it here for simplicity. Once an efficient estimator $\hat{\theta}_{1n}$ of θ_1 is found, method 2 can often be used to construct an efficient estimator of θ_2 .

While no general theorem yet exists, the estimator $\hat{\theta}_{1n}$ defined above (or variations thereon involving suitable smoothing and truncation) has been shown to be asymptotically efficient in several important problems, a notable example being the errors in variables models studied by Bickel and Ritov (1984). Roughly speaking, the fact that \mathbf{i}_1^* is orthogonal to $\mathbf{i}_2, \dots, \mathbf{i}_d$, the scores for θ_2 , permits the use of an inefficient estimator $\bar{\theta}_{2n}$ to estimate out the "nuisance parameter" θ_2 . This should be contrasted with solving (or approximating by a one-step solution)

$$\sum_{i=1}^n \mathbf{i}_1(\theta_1, \bar{\theta}_{2n}) = 0$$

for θ_1 , a method which is known to produce inefficient estimates of θ_1 in general; see e.g. Gong and Samaniego (1981).

The main drawback of the method is that it requires calculation of the efficient score function \mathbf{i}_1^* . Thus the method depends heavily on being able to calculate projections onto $[\mathbf{i}_2] = \dot{P}_{\theta_2} = \dot{P}_C$, which often necessitates calculation of the inverse of the information operator $\mathbf{i}_2^T \mathbf{i}_2 = I_{22}$. When $\mathbf{i}_1^* = \mathbf{i}_1$ so \mathbf{i}_1 is orthogonal to $[\mathbf{i}_2] = \dot{P}_{\theta_2}$, then "adaptation" with respect to $\theta_2 = C$ is possible, and method 1 becomes essentially the method used to construct efficient estimates in this case; see e.g. Stone (1975) and Bickel (1982).

Method 2: Efficient Estimation of θ_2 for known θ_1

Now suppose that an *efficient* estimate $\tilde{\theta}_{2n}$ of θ_2 is available if θ_1 is *known*. We denote this estimator by $\tilde{\theta}_{2n}(\theta_1)$ because it depends on the "known" value of θ_1 . Substitution of this estimate of θ_2 into the ordinary score for θ_1

and a \sqrt{n} -consistent estimator $\bar{\theta}_{1n}$ of $\theta_1 = (\lambda, \psi)$ is given by

$$\bar{\lambda}_n = \frac{\bar{F}_n(T)}{\int_T^\infty (x - T) dF_n(x)} = \text{reciprocal of estimated mean residual life at } T$$

and, since $1 - F(T) = (1 - \psi)(1 - H(T))$ with $H(x) \equiv \int_0^x G(y) \lambda \exp(-\lambda(x - y)) dy = \int_0^x \frac{F(t)}{\psi} \lambda \exp(-\frac{\lambda}{\psi}(x - y)) dy$, $\bar{\psi}_n > 0$ satisfying

$$\frac{\bar{\psi}_n}{1 - \bar{\psi}_n} \bar{F}_n(T) = \int_0^T \exp(-\frac{\bar{\lambda}_n}{\bar{\psi}_n}(T - y)) dF_n(y).$$

Thus method 2 suggests that

$$\hat{\theta}_{1n} = \bar{\theta}_{1n} + \hat{I}_{11}^{-1} \frac{1}{n} \sum_{i=1}^n \hat{i}_1(\bar{\theta}_{1n}, \hat{g}_n(\cdot, \bar{\theta}_{1n}), X_i)$$

with

$$\hat{I}_{11} \equiv \frac{1}{n} \sum_{i=1}^n \hat{i}_1 \hat{i}_1^T,$$

$\hat{i}_1 \equiv \hat{i}_1(\bar{\theta}_{1n}, \hat{g}_n(\cdot, \bar{\theta}_{1n}), X_i)$, will yield an efficient estimate of θ_1 . Substitution of $\hat{\theta}_{1n}$ into (6) should then yield an efficient estimate of G .

4. PROBLEMS

Statistician's have a large, well - stocked tool-box for dealing with classical parametric models, and a growing companion set of tools for handling completely nonparametric models. The choice of tools for dealing with the rich middle ground of semiparametric models is, however, still relatively limited, and the few available tools are not all well suited for the job. Many important problems remain. Here is a partial list:

(a). **Calculation of lower bounds.** If the projection $\Pi(\hat{i}_0 | \hat{P}_G)$ in section 2 can be calculated, then so can the efficient score function \hat{i}_0 , the effective information $I_{11}^*(\theta)$, and the efficient influence curve \hat{i}_1 . In many models this projection is simply a conditional expectation, and hence can be calculated easily; but in other models such as the dependent proportional hazards model of 1.E(b) the projection calculation is apparently intractable. More systematic methods, possibly involving iterative, numerical techniques, are needed.

(b). **Construction of efficient estimates.** Huang (1984) has made a preliminary study of method 1 outlined in section 3, but general results concerning the asymptotic efficiency of methods 1 and 2, or variations thereof involving more smoothing, are still needed. Other methods including minimum Hellinger distance estimates, minimum Kullback - Leibler discrepancy estimators, and maximum - likelihood estimators obtained via EM - algorithms all need further development and sharpening in the context of semiparametric models. Efficient estimates are still unknown for many of the models given in section 1.

(c). **Identifiability and regularity criteria.** For many semiparametric models, further work on identifiability and conditions for regularity of submodels is still needed before work on estimation can get underway. For examples of such studies, see the papers by Heckman and Singer (1984) and Elbers and Ridder (1983) concerning identifiability issues for the models of 1.E(b) and 1.E(c). Classical regularity investigations of translation and parametric models, which carry over to many group models are given by Hájek (1962, 1972).

BIBLIOGRAPHY

- Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A., (1983). "Information and asymptotic efficiency in parametric - nonparametric models," *Ann. Statist.*, vol. 11, pp. 432 - 452.
- Bennett, S., (1983). "Analysis of survival data by the proportional odds model," *Statistics in Medicine*, vol. 2, pp. 273 - 277.
- Beran, R., (1977a). "Estimating a distribution function," *Ann. Statist.*, vol. 5, pp. 400 - 404.
- Beran, R., (1977b). "Robust location estimates," *Ann. Statist.*, vol. 5, pp. 431 - 444.
- Bhattacharya, P. K., Chernoff, H., and Yang, S. S., (1983). "Nonparametric estimation of the slope of a truncated regression," *Ann. Statist.*, vol. 11, pp. 505 - 514.
- Bickel, P. J., (1982). "On adaptive estimation," *Ann. Statist.*, vol. 10, pp. 647 - 671.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A., (1986). "Efficient and adaptive inference in semiparametric models," *Forthcoming monograph*, Johns Hopkins University Press, Baltimore.
- Bickel, P. J. and Lehmann, E. L., (1975). "Descriptive statistics for nonparametric models. I. Introduction.," *Ann. Statist.*, vol. 3, pp. 1038 - 1044.
- Bickel, P. J. and Ritov, Y., (1984). "Efficient estimation in the errors in variables model," *Preprint*, Dept. of Statistics, University of California, Berkeley.
- Branston, D., (1976). "Models of single lane time headway distributions," *Transportation Science*, vol. 10, pp. 125 - 148.
- Breiman, L. and Friedman, J., (1982). "Estimating optimal transformations for multiple regression and correlation," *Preprint*.
- Buckley, J. and James, I., (1979). "Linear regression with censored data," *Biometrika*, vol. 66, pp. 429 - 436.
- Carroll, R. J., (1984). "Adaptation for the slope in simple logistic regression with an intercept in the structural errors in variables model," *Preprint*.
- Carroll, R. J., (1984). "A general technique for computing information bounds in errors in variables structural models," *Preprint*.
- Clayton, D., (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika*, vol. 65, pp. 141 - 151.
- Clayton, D. and Cuzick, J., (1984). "Multivariate generalizations of the proportional hazards model," *Preprint*.
- Clayton, D. and Cuzick, J., (1985). "An approach to inference for rank - regression models with right - censored data," *Preprint*.
- Clayton, D. and Cuzick, J., (1985). "The semi-parametric Pareto model for regression analysis of survival times," *Proceedings of the ISI Meeting*, Preprint.
- Cosslett, S. R., (1981). "Maximum likelihood estimation for choice - based samples," *Econometrica*, vol. 49, pp. 1289 - 1316.

- Jewell, N. P., (1984b). "Least squares estimation of the slope of a truncated regression," *Preprint*, Department of Biostatistics, University of California, Berkeley.
- Kiefer, J. and Wolfowitz, J., (1956). "Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters," *Ann. Math. Statist.*, vol. 27, pp. 887 - 906.
- Koshevnik, Yu. A. and Levit, B. Ya., (1976). "On a nonparametric analogue of the information matrix," *Theor. Prob. Appl.*, vol. 21, pp. 738 - 753.
- Lafier, J. and Kinman, T. D., (1965). "An RR Lyrae star survey with the Lick 20 - inch astrophysical J., *Suppl.*, vol. 11, pp. 216 - 222.
- Laird, N., (1978). "Nonparametric maximum likelihood estimation of a mixing distribution," *J. Amer. Statist. Assoc.*, vol. 73, pp. 805 - 811.
- Lambert, D. and Tierney, L., (1984a). "Asymptotic efficiency of estimators of functionals of mixed distributions," *Ann. Statist.*, vol. 12, pp. 1380 - 1387.
- Lambert, D. and Tierney, L., (1984b). "Asymptotic properties of maximum likelihood estimates in the mixed Poisson model," *Ann. Statist.*, vol. 12, pp. 1388 - 1399.
- Le Cam, L., (1972). "Limits of experiments," *Proc. Sixth Berkeley Symp. Math. Statist. and Prob.*, vol. 1, pp. 245 - 261, University of California Press, Berkeley, California.
- Le Cam, L., (1984). "Review of Ibragimov and Has'minskii (1981) and Pfanzagl (1982)," *Bull. (New Series) Amer. Math. Soc.*, vol. 11, pp. 391 - 400.
- Lehmann, E. L., (1953). "The power of rank tests," *Ann. Math. Statist.*, vol. 24, pp. 23 - 43.
- Lehmann, E. L., (1983). *Theory of Point Estimation*, Wiley, New York.
- Levit, B. Ya., (1978). "Infinite - dimensional informational lower bounds," *Theor. Prob. Applic.*, vol. 20, pp. 723 - 740.
- Lindsay, B., (1980). "Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators," *Philos. Trans. Roy. Soc. London Ser A*, vol. 296, pp. 639 - 665.
- Lindsay, B., (1983a). "The geometry of mixture likelihoods, part I," *Ann. Statist.*, vol. 11, pp. 86 - 94.
- Lindsay, B., (1983b). "The geometry of mixture likelihoods, part II," *Ann. Statist.*, vol. 11, pp. 783 - 792.
- McDonald, J., (1983). "Periodic smoothing of time series," *Project Orion Technical Report*, vol. 017, Department of Statistics, Stanford University, Stanford, California.
- Millar, P. W., (1979). "Asymptotic minimax theorems for the sample distribution," *Z. Wahrsch. verw. Gebiete*, vol. 48, pp. 233 - 252.
- Millar, P. W., (1983). "The minimax principle in asymptotic statistical theory," *Proc. Ecole d'Ete St. Flour, Lecture Notes in Math.*, vol. 976, pp. 75 - 265, Springer - Verlag, Berlin.
- Millar, P. W., (1985). "Nonparametric applications of an infinite dimensional convolution theorem," *Z. Wahrsch. verw. Gebiete*, vol. 68, pp. 545 - 556.
- Miller, R. and Halpern, J., (1982). "Regression with censored data," *Biometrika*, vol. 69, pp. 521 - 531.