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 \( m \))

\[
p(x; \theta) = c(\theta)\exp\left(\sum_{i=1}^{k} q_i(\theta)T_i(x)\right)m(x)
\]

for \( \theta \in \Theta \subseteq \mathbb{R}^k, \ x \in X \subseteq \mathbb{R}^d \). 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 \( Y \equiv G = P_\theta \), a fixed distribution on \( X \), and let \( V \) denote a group of (one to one) transformations on \( X \) parametrized by \( \theta \in \Theta \subseteq \mathbb{R}^k \). If \( \nu_\theta \in V \), let \( X = \nu_\theta(Y) \equiv P_\theta \) for \( \theta \in \Theta \).

Examples:

(a) Location. \( X = \mathbb{R}^d, \nu_\theta(x) = x + \theta \) with \( \theta \in \mathbb{R}^d \), and \( P_\theta = P_{\theta'}(-\theta) \).

(b) Elliptic distributions. \( X = \mathbb{R}^d, \nu_\theta(x) = \theta^{-1/2}x \) where \( \theta \) is positive definite and symmetric; \( G = P_\theta \) is spherically symmetric on \( \mathbb{R}^d \). Then \( P = \{P_\theta: \theta \in \Theta\} \) is the \( P_\theta \)-family of elliptic distributions.

(c) Two-sample models. \( X = \mathbb{R}^d, \nu_\theta(x) = \theta_0 + \theta_1 \) where \( \nu_\theta \) is a group of matrices on \( \mathbb{R}^d \), \( \theta = (\mu, \nu) \in \Theta_0 \times \Theta_1 = \Theta \), \( Y = (W, Z) \) with \( W, Z \equiv P_0 \) independent, and \( X = (\nu_\mu(W), \nu_\nu(\nu(W))) \).

(2). By letting the distribution \( P_\theta \) in (1) range over some large class of probability distributions \( \mathbb{G} \) small enough to still allow identification of \( \theta \), or at least some important functions of \( \theta \), yields a semiparametric model

\[
P = \{P_{\theta,G}: \theta \in \Theta, G \in \mathbb{G}\}.
\]

Examples:

(a) If \( X = \mathbb{R}^1 \) in 1(a) above and \( G \) is the family of distributions symmetric about 0, \( P \) is the classical symmetric location family.

(b) If \( X \) and \( \Theta \) are as in 1(b) above and \( G \) is the family of all spherical symmetric distributions, then \( P \) is the family of all elliptical 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
   \[ P = \{ P_{\theta, y}: \theta \in \Theta, \gamma \in G \} \]
   where \( \theta \) is the finite dimensional Euclidean component of the model and \( \gamma \) is the nonparametric or infinite-dimensional component of the basic model.

2. Suppose that \( Z \equiv H \) on \( R^d \).

3. Given \( Z = z \), replace \( \theta \) (or a component thereof) in the basic model by a semiparametric regression function \( r(\gamma, z) \) taking values in \( \Theta \) where \( \gamma \in \Gamma \subset \text{some} R^k \). 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: \( r(\gamma, z) = \gamma_1 z_1 + r(z_2) \), where \( z = (z_1, z_2) \), and \( r \) is smooth.
   (e) Projection pursuit: \( r(\gamma, z) = r(\gamma z) \) where \( |\gamma| = 1 \) and \( r: R^1 \to R^1 \) is smooth.
   (f) Signal-noise: \( r(\gamma z) \) where \( r: R^1 \to R^1 \) is periodic with period 1 so that \( \gamma \) 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 B(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.
and Oakes (1982), and has been generalized by Gill (1984). Related regression models are discussed by Ridder and Verbeke (1983) and Elbers and Ridder (1983).

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

$$Y_1 = Z + \epsilon_1$$

$$Y_2 = \alpha + \beta Z + \epsilon_2$$

where $Z \equiv H$ (non-Gaussian) and $\epsilon \equiv (\epsilon_1, \epsilon_2) \equiv N(0, \Sigma)$. The resulting mixture model is an errors in variables regression model. Consistent maximum likelihood estimates were obtained by Kiefer and Woldowitz (1956), but lower bounds for estimation of $(\alpha, \beta)$ together with asymptotically efficient estimates attaining the bounds were first obtained by Elbers and Ridder (1984).

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

$$P_H(Y \geq y) = \int_0^y \exp(-z) \, dz.$$ 

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'minekili and Ibragimov (1983).

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

Examples:

a. Mixing. The mixing models of $E$ are censoring models with $X^* = T(Y, Z) = Y.$

b. Random right censorship. In this type of censoring, which has received much use in survival analysis, $X^* = (X_1', X_2') = T(X_1, X_2) = (X_1 \wedge X_2, 1_{\{X_1 \neq X_2\}})$. Random right censoring meshes extremely well with Cox’s proportional hazards regression model as discussed in $D$($s$). 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 (1994) 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 $X^* = T(X_1, X_2) = X_1 + X_2$ where $X_1$ and $X_2$ are independent. The traffic model of Branson (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 $X \equiv P_{\theta, \varphi} \in P$, a semiparametric model. Then suppose that $K_i(z), i = 1, \ldots, s$ is a collection of known non-negative biasing kernels and that $\lambda_i, i = 1, \ldots, s$ is a probability distribution on $[1, \ldots, s]$. Then the biased sampling distribution corresponding to $P_{\theta, \varphi}$.
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 √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 ≥ 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 (Hajek - Le Cam) bounds for estimation in 'regular' models. The Hajek - 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), Kolesnikov and Levit (1978), Levit (1978), Millar (1979,1983,1985) Pfanzagl (1962), 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 \( X_1, \ldots, X_n \) are iid \( P_\theta \in \mathcal{P} = \{ P_\theta: \theta \in \Theta \} \) where \( \Theta \subseteq \mathbb{R}^d \) is open, that \( \mathcal{P} \) is dominated by a (sigma - finite) measure \( \mu \) on \( \Theta \), and let \( p(\cdot, \theta) = \frac{dP_\theta}{d\mu} \) for \( \theta \in \Theta \). Then the classical log-likelihood of an observation \( X \) is
\[
\ell(\theta, X) = \log p(X, \theta).
\]

the scores vector \( \ell \) is
\[
\ell(\theta, X) = \nabla \ell(\theta, X) = \frac{1}{p(X, \theta)} \left( \frac{\partial}{\partial \theta_1} p(X, \theta), \ldots, \frac{\partial}{\partial \theta_d} p(X, \theta) \right)^T,
\]

and the Fisher information matrix for \( \theta \) is
\[
I(\theta) = E_\theta \left[ \ell(\theta, X) \ell(\theta, X)^T \right].
\]

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 \( \mathbb{R}^d \), is:

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

If we focus on estimation of the first component \( \hat{\theta}_1 \in \mathbb{R}^1 \) of \( \theta \), it follows immediately from (1), the definition of \( I(\theta) \), and standard \( L_q \) projection or regression theory that

\[
n \text{Var}_E[\hat{\theta}_1] \geq \sup_{b \in \mathbb{R}^d} \frac{b_1^2}{b^T I(\theta)^{-1} b} = I_{11}(\theta) \tag{2}
\]

\[
= \inf_{c \in \mathbb{R}^d, c_1 = 1} \frac{1}{E_\theta \left[ (\hat{\theta}_1 - c_2 \hat{\theta}_2 - \cdots - c_d \hat{\theta}_d)^2 \right]}
\]

\[
= \frac{1}{I_{11}(\theta) - I_{12}(\theta) I_{22}(\theta)^{-1} I_{21}(\theta)} = \frac{1}{I_{11}(\theta)}
\]

23.1-20
\[ l_n(\theta_n) - l_n(\theta) = \log \left[ \frac{p_n(X, \theta_n)}{p_n(X, \theta)} \right]. \]

then \( P = \{ P_{n, \theta} : \theta \in \Theta \} \) is locally asymptotically normal (LAN) at \( \theta \) if there is a vector of \( L_2(P_\theta) \) functions \( b_n(\theta) \) and a nonsingular matrix \( I(\theta) \) such that, with
\[ l_n(\theta_n) - l_n(\theta) = b_n(\theta)^T h - \frac{1}{2} h^T I(\theta) h + R_n(\theta, h), \] (9)

it follows that, in \( P_{n, \theta} \cdot \) probability,
(i) \( R_n(\theta, h) \rightarrow_p 0 \) uniformly on bounded \( h \)-sets, and
(ii) \( b_n(\theta) \rightarrow_d N(0, I(\theta)). \)

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) \( b_n(\theta) = n^{-1/2} \sum_{i=1}^n \tilde{t}(\theta, X_i) \) where \( \tilde{t} \) is the scores vector (for \( n = 1 \)) and \( I(\theta) \) is the information matrix.

Because of our interest here in the parametric component \( \theta \) of a semiparametric model \( P = \{ P_{n, \theta} \} \), we formulate versions of the convolution and asymptotic minimax bounds for the first component \( \theta_1 \) of \( \theta \).

A sequence of estimators \( T_{1n} \) of \( \theta_1 \) is regular at \( \theta \) if, under \( P_{\theta_0} \)
\[ \sqrt{n} (T_{1n} - \theta_{1n}) \rightarrow_d T_1 \] for every \( \theta_{1n} = \theta + n^{-1/2} h \) where the distribution \( L(T_1) \) of \( T_1 \) does not depend on \( h \).

**Theorem 1.** (Hájeck, 1970). Suppose that \( P \) is LAN at \( \theta \) and that \( T_{1n} \) is a regular estimator with limit distribution \( L(T_1) \). Then
\[ T_1 \cong Z_1 + W_1 \] (9)
where \( Z_1 \cong N(0, I_{11}(\theta)) \), \( I_{11}(\theta) \) is as in (3), and \( W_1 \) is independent of \( Z_1 \).

Thus any regular estimator \( T_{1n} \) of \( \theta_1 \) must have a limit distribution which is at least as dispersed as \( N(0, I_{11}(\theta)) \), 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 \( \omega: \mathbb{R}^1 \rightarrow \mathbb{R}^+ \) satisfies:
(i) \( \omega(z) = \omega(-z) \) for all \( z \in \mathbb{R} \),
(ii) \( \omega(0) = 0 \), \( \omega(z) \) increases in \( z \geq 0 \),
(iii) \( E \omega(\sigma Z) < \infty \) for all \( \sigma > 0 \) where \( Z \cong N(0,1) \).

**Theorem 2.** (Hájeck, 1972). Suppose that \( P \) is LAN at \( \theta \) and that \( \omega \) satisfies (i) - (iii). Then, for any estimator \( T_{1n} \) of \( \theta_1 \),
\[ \liminf_{n \rightarrow \infty} \sup_{\theta_0, \theta_1} E_{\theta_0, \theta_1} \omega(\sqrt{n} (T_{1n} - \theta_{1n})) \geq E \omega(Z_1) \] (10)
where \( Z_1 \cong N(0, I_{11}(\theta)) \) 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 Hájeck - 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, \( i_1^* \) for \( \theta_1 \),

\[
i_1^* = i_1 - \frac{1}{2} \frac{\partial^2}{\partial \theta_1^2} \sum_{i=1}^n \bar{S}_i \frac{\partial}{\partial \theta_1} \bar{S}_i = i_1 - \bar{E}(i_1 | \bar{S}_n)
\]

and the effective information

\[
I_{11}^*(\theta) = E(i_1^* i_1^*).
\]

Furthermore, suppose that \( \overline{\theta}_n \) is a \( \sqrt{n} \) consistent estimator of \( \theta \).

\[
\sqrt{n} \overline{\theta}_n - \theta = O_p(1).
\]

Then define \( \overline{\theta}_m \) to be either a solution of the efficient score equation

\[
\sum_{i=1}^n i_1^*(\overline{\theta}_m, \overline{\theta}_m, X_i) = 0,
\]

or a one-step approximation thereof:

\[
\overline{\theta}_m = \overline{\theta}_m + \frac{1}{n} \sum_{i=1}^n \frac{i_1^*(\overline{\theta}_m, X_i)}{I_{11}^*(\overline{\theta}_m)}
\]

(1)

where \( \overline{I}_1 \) is the efficient influence curve for \( \theta_1 \). Additional smoothing may also be required in forming the sums in (1), but we have omitted it here for simplicity. Once an efficient estimator \( \overline{\theta}_m \) of \( \theta_1 \) is found, method 2 can often be used to construct an efficient estimator of \( \theta_2 \).

While no general theorem yet exists, the estimator \( \overline{\theta}_m \) defined above (or variations thereof 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 \( \overline{I}_1^* \) is orthogonal to \( \overline{I}_2, \overline{I}_3, \ldots, \overline{I}_k \), the scores for \( \theta_2 \), permits the use of an inefficient estimator \( \overline{\theta}_m \) to estimate out the "nuisance parameter" \( \theta_2 \). This should be contrasted with solving (or approximating by a one-step solution)

\[
\sum_{i=1}^n i_1^*(\theta_1, \overline{\theta}_m) = 0
\]

for \( \theta_1 \), a method which is known to produce inefficient estimates of \( \theta_1 \) 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 \( \overline{I}_1^* \). Thus the method depends heavily on being able to calculate projections onto \( \overline{[i_2]} = \overline{P}_2 = \overline{P}_2 \), which often necessitates calculation of the inverse of the information operator \( \overline{I}_1^* i_2^* = I_{22} \). When \( \overline{I}_1^* = \overline{I}_1 \), so \( i_1 \) is orthogonal to \( \overline{[i_2]} = \overline{P}_2 \), then "adaptation" with respect to \( \theta_2 = 0 \) 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 \( \theta_2 \) for known \( \theta_1 \)

Now suppose that an efficient estimate \( \overline{\theta}_m \) of \( \theta_2 \) is available if \( \theta_1 \) is known. We denote this estimator by \( \overline{\theta}_m(\theta_1) \) because it depends on the "known" value of \( \theta_1 \). Substitution of this estimate of \( \theta_2 \) into the ordinary score for \( \theta_1 \)
and a \( \sqrt{n} \) - consistent estimator \( \hat{\theta}_{1n} \) of \( \theta_1 = (\lambda, \psi) \) is given by
\[
\hat{\lambda}_n = \frac{\hat{F}_n(T)}{\int_T^\infty (x - T) dF_n(x)} = \text{reciprocal of estimated mean residual life at } T
\]
and,
\[
1 - F(T) = (1 - \psi)(1 - \hat{H}(T)) \quad \text{with } \quad \hat{H}(x) = \int_0^x G(y) \lambda \exp(-\lambda(x-y)) dy = \int_0^x \tilde{F}(t) \lambda \exp(-\lambda \frac{\psi}{\psi}(x-y)) dy.
\]
\( \hat{\psi}_n > 0 \) satisfying
\[
\frac{\hat{\psi}_n}{1 - \hat{\psi}_n} \hat{F}_n(T) = \int_0^T \exp\left(-\frac{\hat{\psi}_n}{\hat{\psi}_n} (T-y)\right) dF_n(y).
\]
Thus method 2 suggests that
\[
\hat{\theta}_{1n} = \hat{\theta}_{1n} + \hat{r}_{11}^{-1} \frac{1}{n} \sum_{i=1}^n \hat{1}_i(\hat{\theta}_{1n}, \hat{\theta}_n, \lambda, \psi, \eta, \xi_i)
\]
with
\[
\hat{r}_{11} = \frac{1}{n} \sum_{i=1}^n \hat{1}_i \hat{1}_i^T.
\]
\( \hat{1}_i = \hat{1}_i(\hat{\theta}_{1n}, \hat{\theta}_n, \lambda, \psi, \eta, \xi_i), \) will yield an efficient estimate of \( \theta_1 \). 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 \( \Pi(\hat{\theta}, \hat{\theta}_q) \) in section 2 can be calculated, then so can the efficient score function \( \hat{b}_n \), the effective information \( \hat{r}_{11}(\theta) \), 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 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 Ebers 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


