CHAPTER 1

Semiparametric Models: a Review of Progress since BKRW (1993)

Jon A. Wellner, Chris A. J. Klaassen, Ya'acov Ritov

University of Washington
Statistics
Box 354322
Seattle, Washington 98195-4322
U.S.A.
E-mail: jaw@stat.washington.edu

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
The Netherlands
E-mail: chrisk@science.uva.nl

Department of Statistics
The Hebrew University of Jerusalem
Mt. Scopus, Jerusalem 91905
Israel
E-mail: yaacov@mscc.huji.ac.il

This paper sketches a review of the developments in semiparametric statistics since the publication in 1993 of the monograph by Bickel, Klaassen, Ritov, and Wellner.

1. Introduction

This paper gives a brief review of some of the major theoretical developments in the theory of semiparametric models since publication of our jointly authored book, Bickel, Klaassen, Ritov, and Wellner (1993), henceforth referred to as BKRW (1993). It is, for reasons of space, a very selective and somewhat personal review. We apologize in advance to all of those whose works we have not covered for any reason.

A major focus in semiparametric theory is on asymptotic efficiency. A special case of this semiparametric efficiency occurs when the least favorable parametric submodel of the semiparametric model is a natural parametric model. Typically this natural parametric model is the model from which the semiparametric model is built by relaxing distributional assumptions. For example in the semiparametric symmetric location model the least favorable parametric submodel is the symmetric location model with the symmetric density known. In the semiparametric linear regression model with the error distribution unknown with mean zero, the least favorable parametric submodel is the linear regression model with the error distribution known. The first semiparametrically efficient procedures were developed in these cases and they were called adaptive because they adapted themselves to the unknown underlying density. Stone (1975) and Beran (1974) were the first to construct efficient estimators in the semiparametric symmetric location model, preceded by van Eeden (1970) who proved efficiency of her estimator with the symmetric densities assumed to be strongly unimodal. A milestone in this line of research was the paper by Bickel (1982). Actually, existence of adaptive procedures was suggested already in the fifties by Stein (1956) and the first adaptive test was presented by Hájek (1962).

Meanwhile Cox (1972) introduced the proportional hazards model and his estimators for the parametric part of the model, and Cox (1975) introduced his notion of "partial likelihood". There was a flurry of work in the late 1970's in an effort to understand the efficiency properties of Cox's "partial likelihood" estimators: e.g. Efron (1977), Oakes (1977), and Kay (1979). The work of the first author of the present paper (culminating in Begun, Hall, Huang and Wellner (1983)), began with an effort to rework and generalize the calculations of Efron (1977) along the lines of some of the modern information bound theory in papers of Beran (1977a, 1977b).

The first appearances of the term "semiparametric" in the literature (of which we are aware) occur in a *Biometrics* paper by Gail, Santner, and Brown (1980) and in a paper in *Demography* by Finnas and Hoem (1980). Within a year the term was also used by Oakes (1981) in his influential review of work on the Cox (1972) model, and by Turnbull (1981) in his *Mathematical Reviews* review of Kalbfleisch and Prentice (1980). Subsequently the term was applied by Epstein (1982, 1983) in reviews of Whitehead (1980) (who used the terminology "partially parametric") and Oakes (1981), by Louis, Mosteller, and McPeek (1982), page 95, and by Andersen (1982), page 67. Unfortunately, the first author of the present paper in writing Begun, Hall, Huang, and Wellner (1983) used the terminology "parametric - nonparametric" and "mixed model". The terminology "semiparametric" became accepted, however, as can be seen from Figure 1 which shows appearance of the term "semiparametric" in title, keywords, or abstract in three major indexes of statistical literature: MathSciNet, the Current Index of Statistics, and the ISI Web of Science.

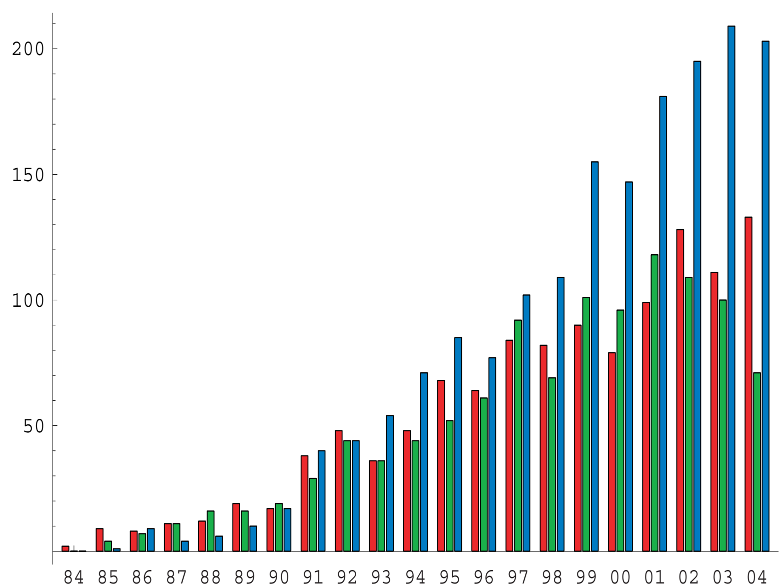

Fig. 1. Numbers of papers with "semiparametric" in title, keywords, or abstract, by year, 1984 - 2004. Red = MathSciNet; Green = Current Index of Statistics (CIS); Blue = ISI Web of Science

2. Progress in semiparametric model theory since 1993: a short review

The theory of estimation and testing for semiparametric models has been developing rapidly since the publication of BKRW in (1993). Here we briefly survey some of the most important of these developments from our perspective and pose some questions and challenges for the future. We will not attempt to review the many applications of semiparametric models since they have become too numerous to review in the limited space available here. [A search on MathSciNet in early May 2005 for "semiparametric" gave 1185 hits.] Our review will be broken down according to the following (somewhat arbitrary and overlapping) categories:

- Missing data models.
- Testing and profile likelihood theory.
- Semiparametric mixture model theory.
- Rates of convergence via empirical process methods.
- Bayes methods and theory.
- Model selection methods.
- Empirical likelihood.
- Transformation and frailty models.
- Semiparametric regression models.
- Extensions to non-i.i.d. data.
- Critiques and possible alternative theories.

2.1. *Missing data models*

[A search on MathSciNet in early May 2005 for "semiparametric" and "missing data" gave 15 hits.] A major development in this area was the systematic development of information bounds for semiparametric regression models with covariates missing at random by Robins, Rotnitzky, and Zhao (1994), Robins, Rotnitzky, and Zhao (1995), and Robins, Hsieh, and Newey (1995); see also Robins and Rotnitzky (1992), Robins and Rotnitzky (1995), and Nan, Emond, and Wellner (2004). For another recent treatment of the information bound calculations under Coarsening At Random (CAR) and Missing At Random (MAR), see van der Vaart (1998), pages 379 - 383.

The information bounds for missing data models have been shown to be achievable in some special cases: for examples involving two-stage sampling, see Breslow, McNeney, and Wellner (2003), and see Chen (2002), Chen (2004), and Wang, Linton, and Härdle (2004) for further examples and recent developments. Much further work is needed in this area.

2.2. *Testing and profile likelihood theory*

[A search on MathSciNet in early May 2005 for "semiparametric" and "profile likelihood" gave 13 hits; a search for "semiparametric" and "testing" gave 233 hits.] Although BKRW (1993) did not manage to treat the theory of tests for semiparametric models, the literature developed rapidly in this area during the mid and late 1990's, with contributions by Choi, Hall, and Schick (1996), Murphy and van der Vaart (1997), and Murphy and van der Vaart (2000). In particular, Severini and Wong (1991) initiated the study of profile likelihood methods in semiparametric settings, and their study was developed further by Murphy and van der Vaart (1997) and Murphy and van der Vaart (2000). Murphy and van der Vaart (1997, 2000) show that semiparametric profile likelihoods have quadratic expansions in the efficient scores under appropriate Donsker type hypotheses on the scores corresponding to a least favorable sub-model and a certain "no-bias" condition. This important development opens the door to likelihood ratio type tests and confidence intervals in many semiparametric models for which the least favorable sub-models can be constructed. The main difficulty in applying the results of Murphy and van der Vaart (2000) seems often to be in construction of least favorable submodels with the right properties. Severini and Staniswalis (1994) develop methods based on combining techniques from profile likelihood and quasi-likelihood considerations, while Lee, Kosorok, and Fine (2005) propose Bayesian MCMC methods applied to the semiparametric profile likelihood. Banerjee (2005) has studied the power behavior of likelihood ratio tests under contiguous alternatives, and shows that the limiting distributions under local alternatives are non-central chi-square with shift parameter involving a quadratic form in the efficient information matrix. Murphy and Van der Vaart (1999) study the use of "observed information" in semiparametric models and applications thereof to testing.

Testing a parametric fixed link single-index regression model against a semiparametric alternative single-index model was considered by Horowitz and Härdle (1994), but with the parameter θ_0 involved in the single index under both null and alternative. The case of index parameter allowed to differ under the alternative was considered by Härdle, Spokoiny, and Sperlich (1997). Kauermann and Tutz (2001) consider testing certain classes of parametric and semiparametric regression models against general smooth alternatives. On the other hand, Bickel, Ritov, and Stoker (2005a) argue that there is no real notion of optimality in testing of semiparametric composite hypotheses. Any test would have negligible power against departures in most directions. Their recommendation is to use tailor-made tests, which concentrate power in directions which are important to the investigator. These general ideas were applied to index models in Bickel, Ritov, and Stoker (2005b).

2.3. *Semiparametric mixture model theory*

[A search on MathSciNet in early May 2005 for "semiparametric" and "mixture model" gave 37 hits.] In a classical paper proposing models alternative to those considered by Neyman and Scott (1948), Kiefer and Wolfowitz (1956) showed that maximum likelihood estimators are consistent in a large class of semiparametric mixture models before the term "semiparametric" was in existence. Although other less satisfactory estimators had been constructed for many models of this type during the 1970's and 1980's (see e.g. van der Vaart (1988)), efficiency and asymptotic normality of maximum likelihood estimators were completely unknown through the mid-1990's. But van der Vaart (1996) succeeded in using empirical process theory methods together with methods and results of Pfanzagl (1988) and Pfanzagl (1990) to establish asymptotic normality and efficiency of the maximum likelihood estimators for several important examples of this type of model (including an exponential frailty model, a normal theory errors in variables model, and a model involving scale mixtures of Gaussians). It seems difficult to formulate a completely satisfactory general theorem, but it also seems clear that the methods of van der Vaart (1996) will apply to a wide range of semiparametric mixture models.

2.4. *Rates of convergence via empirical process methods*

[A search on MathSciNet in early may 2005 for "semiparametric" and "convergence rate" gave 27 hits. Searching for "nonparametric" and "convergence rate" gave 214 hits.] Rates of convergence of minimum contrast estimators, maximum likelihood estimators, and variants of maximum likelihood involving sieves and penalization, mostly aimed at nonparametric settings, were a topic of considerable research during the 1990's, beginning with Wong and Severini (1991), Birgé and Massart (1993), Shen and Wong (1994), and Wong and Shen (1995). The results of these authors rely on sharp bounds for local oscillations of empirical processes indexed by the classes of functions involved in the maximization, and hence are closely related to the available bounds for suprema of empirical processes. See van der Vaart and Wellner (1996) sections 3.2 and 3.4 for a recasting of those results.

This initial progress continued with van de Geer (1996), Birgé and Massart (1998), Shen (1997), and Shen and He (1997).

The results in these works have important consequences for maximum likelihood estimators as well as sieved and penalized maximum likelihood estimators in semiparametric models. For example, Huang (1996) used the methods of Birgé and Massart (1993) and Wong and Shen (1995) to obtain rates of convergence of maximum likelihood estimators for Cox's proportional hazards model with current status data. See van der Vaart (2002), section 8, pages 424 - 432, for a summary of the methods and an alternative treatment of Huang's results. van der Vaart (2002) also gives a number of other nice applications of empirical process theory to problems in semiparametric models.

2.5. *Bayes methods and theory*

[A search on MathSciNet in early May 2005 for "semiparametric" and "Bayesian" gave 121 hits.] Bayes estimators and procedures have been proposed for a wide range of semiparametric models: see Lenk (1999) for a Bayesian approach to semiparametric regression, Müller and Roeder (1997) for a Bayesian model for case-control studies, Vidakovic (1998) for Bayes methods in connnection with wavelet based nonparametric estimation, Newton, Czado, and Chappell (1996) for Bayes inference for semiparametric binary regression, and Ghosal and van der Vaart (2001) for Bayes estimation with mixtures of normal densities. Lazar (2003) gives an interesting Bayes approach to empirical likelihood (see below), while Sinha, Ibrahim, and Chen (2003) give an interesting Bayesian justification of Cox's partial likelihood. Much of the popularity of Bayes methods is due to the new computational tools available; see e.g. Gilks, Richardson, and Speigelhalter (1999), Carlin and Louis (2000), and Robert and Casella (2004). Considerable progress has been made in understanding consistency issues and rates of convergence of Bayes procedures (see e.g. Ghosal, Ghosh, and Samanta (1995), Ghosal, Ghosh, and van der Vaart (2000), Shen and Wasserman (2001), and Huang (2004)), but major gaps remain in the theory. For example, a suitably general Bernstein - von Mises theorem is still lacking despite several initial efforts in this direction. Freedman (1999) gives negative results for non-smooth functionals in the context of nonparametric regression (which are not surprising from the perspective of semiparametric information bounds), while Kim and Lee (2004) give more positive results in a right censoring model with smooth functionals, and Shen (2002) gives some preliminary general results. Kleijn and van der Vaart (2002) give a treatment of Bayes estimators in the case of miss-specified models. From the examples, it is clear that Bayes procedures need not even be consistent in general and that care is needed with respect to the choice of priors. The growing number of examples and special cases point to the need for a more complete theoretical understanding.

2.6. *Model selection methods*

[A search on MathSciNet in early May 2005 for "semiparametric" and "model selection" gave 28 hits.] Theoretical understanding of model selection methods in nonparametric estimation problems has progressed rapidly during the last 10 years with major contributions by Birgé and Massart (1997), Birgé and Massart (1998), Barron, Birgé, and Massart (1999), and Birgé and Massart (2001). Also see Massart (2000) and Massart's forthcoming St. Flour Lecture notes from 2003. These developments have begun to have some impact on semiparametric estimation as well: Raftery, Madigan, and Volinsky (1996) and Raftery, Madigan, Volinsky, and Kronmal (1997) used Bayesian methods for covariate selection in Cox models. Tibshirani (1997) introduces "lasso" methods for proportional hazards models (which involve an L_1 penalty term), while Fan and Li (2002) propose an alternative approach based on a non-concave penalty, and extend these methods to a class of frailty models (see below). Bunea (2004) studies the effect of covariate selection methods on inference in a partly linear regression model, and this is carried over to non-proportional hazards models in survival analysis by Bunea and McKeague (2005). It would be of some interest to extend these developments to the models considered by Huang (1999). We suspect that much more remains to be done to fully understand the advantages and disadvantages of various model-selection strategies.

2.7. *Empirical likelihood*

[A search on MathSciNet in early May 2005 for "semiparametric" and "empirical likelihood" gave 18 hits; searching for "empirical likelihood" alone gave 192 hits.] Owen (1988) and Owen (1990) introduced the notion of empirical likelihood and showed how a reasonable facsimile of the standard theorem for the likelihood ratio statistic in regular parametric models continues to hold for finite-dimensional smooth functionals $\nu(P)$, $\nu : \mathcal{P} \rightarrow \mathbb{R}^d$. The basic notion in Owen's theory involves estimation of P in the restricted model $\mathcal{P}_0 = \{P \in \mathcal{M} : \ \nu(P) = t_0\}$ where $t_0 \in \mathbb{R}^d$ is fixed. The resulting model can be viewed as a semiparametric model with a tangent space having finite co-dimension in $L_2^0(P)$, one of the topics treated in BKRW (see section 6.2, pages 222- 229) and much earlier by Koshevnik and Levit (1976). This led to a considerable development of "empirical likelihood" based methods in connection with estimating equation approaches to a wide variety of semiparametric models: see e.g. Owen (1991), Qin (1993), Qin and Lawless (1994), Qin and Lawless (1995), and Lazar and Mykland (1999). Qin (1998) and Zhou, Fine, and Yandell (2002) give applications to mixture models.

2.8. *Transformation and frailty models*

[A search on MathSciNet in early May 2005 for "semiparametric" and "frailty" gave 28 hits searching for "semiparametric" and "transformation" gave 35 hits.] Cox (1972) introduced the proportional hazards model, resulting in one of the most cited papers in statistics ever. The Cox proportional hazards model for survival data is a transformation model with the baseline cumulative hazard function as unknown transformation. It is one of the prime examples of a semiparametric model in BKRW. Clayton and Cuzick (1985) generalized the Cox model by introducing frailty as an unobservable random factor in the hazard function. The results of Clayton and Cuzick inspired the key theoretical development in Bickel (1986); also see Klaassen (2005) for further discussion of the differential equations determining the efficient score functions in transformation models. These developments started a stream of papers that propose pragmatic methods in ever more complicated frailty models incorporating e.g. cluster-frailty, censoring etc. Efficiency of the proposed inference procedures is not necessarily a goal here; e.g. Li and Ryan (2002). Murphy (1995) provided key asymptotic theory for maximum likelihood estimation methods in a basic gamma-frailty model, and her methods were extended to more complicated frailty models by Parner (1998).

Another fundamental paper is Bickel and Ritov (1997) which discusses the issue of loss of efficiency when going from core to transformation model. The methods advocated by Bickel and Ritov have been implemented in the case of the binormal ROC model by Zou and Hall (2000), Zou and Hall (2002); see Hsieh and Turnbull (1996) for background material concerning ROC curve estimation. Cai and Moskowitz (2004) give a profile likelihood approach to estimation in ROC models that deserves further study and evaluation.

Copula models for joint distributions have started receiving increasing interest in econometrics, finance, and other application areas. Klaassen and Wellner (1997) identified normal location-scale families as the least favorable sub-models for the class of bivariate normal copula models and constructed efficient estimators. Progress on efficient estimation for other copula models remains as a challenging open problem (with potential recent progress by Chen, Fan, and Tsyrennikov (2004)).

A study of transformation and other semiparametric models with a focus on applications in econometrics is given in Horowitz (1998). A Bayesian approach is presented by Mallick and Walker (2003).

2.9. *Semiparametric regression models*

[A search on MathSciNet in early May 2005 for "semiparametric" and "regression" gave 652 hits.] A motivating class of models for BKRW were the "partially linear semiparametric regression models" studied in section 4.3, pages 107 - 112, BKRW (1993). It was already shown in Ritov and Bickel (1990) that efficient estimators cannot be constructed in these semiparametric models without smoothness assumptions on the class of functions allowed (also see section 2.11 below). On the other hand, Schick (1993) gave a treatment of information bounds and construction of efficient estimators for the parametric component for a general class of models encompassing those treated in BKRW that improved on the results of Cuzick (1992). His approach was via estimation of the efficient influence function.

In the meantime information bounds and efficient estimators have now been constructed for many generalizations and variants of these models. For example, Sasieni (1992a) and Sasieni (1992b) calculated information bounds for a partially linear version of the Cox proportional hazards model. Estimates achieving the bounds were constructed by Huang (1999). Efficient estimation in a different but related class of models was studied by Nielsen, Linton, and Bickel (1998).

Interesting classes of semilinear semiparametric regression models for applications to micro-array data have recently been introduced by Fan, Peng, and Huang (2005) and Huang, Wang, and Zhang (2005). There seem to be a large number of open questions connected with these models and their applications.

For a Bayesian approach to semiparametric regression models, see Seifu, Severini, and Tanner (1999).

2.10. *Extensions to non-i.i.d. data*

[A search on MathSciNet in early May 2005 for "semiparametric" and "time series" gave 139 hits.] Semiparametric theory for times series models was already well underway by the time BKRW was published in (1993). For example, Kreiss (1987) had considered adaptive estimation (in the sense of Stein (1956) and BKRW) for stationary (causal) ARMA processes, while Gassiat (1990), (1993) showed that adaptation is not possible in the case of non-causal ARMA models. Drost, Klaassen, and Werker (1997) and Drost and

Klaassen (1997) generalized the results of Kreiss and others (notably Koul and Schick (1997)) to classes of non-linear time series models. Koul and Schick (1996) considered an interesting random coefficient autoregressive model. The current state of the art is summarized in Greenwood, Müller, and Wefelmeyer (2004).

Consideration of information bounds and efficient estimation more generally for a wide variety of Markov chains and other Markov processes finally emerged in the mid- and late 1990's: see e.g. Greenwood and Wefelmeyer (1995), Schick and Wefelmeyer (1999), and Kessler, Schick, and Wefelmeyer (2001). Bickel and Kwon (2001) (following up on Bickel (1993)) reformulated much of this work and provided considerable unification. Also see the discussion piece by Greenwood, Schick, and Wefelmeyer following Bickel and Kwon (2001).

Another interesting direction of generalization concerns relaxing from "structural" (or i.i.d.) modeling to "functional" (non - i.i.d.) modeling. For an interesting study of the classical normal theory "errors in variables" model under "functional" or "incidental nuisance parameters", see Murphy and van der Vaart (1996). McNeney and Wellner (2000) give a review of information bounds for functional models.

2.11. *Critiques and possible alternative theories*

In two key papers Bickel and Ritov (1988) and Ritov and Bickel (1990) pointed out that attainment of information bounds in semiparametric and nonparametric situations requires additional assumptions on the dimensionality of the parameter space. They gave several explicit examples of differentiable functionals, for which the information bounds are finite and yet are not attained in general by any estimator. One of these examples is the functional $\nu(P) = \int_0^1 p^2(x)dx$ for probability measures P on $[0,1]$ with density p with respect to Lebesgue measure. Another example involves estimation of θ in the partly linear regression model $Y = \theta^T Z + r(X) + \epsilon$ based on observation of (Y, Z, X). In both examples, the standard semiparametric bounds are attained when the parameters p and r are assumed to be smooth enough. However, for other smoothness classes, one can show that the (attained) minimax rate is much slower than $n^{-1/2}$. Moreover, in general, there is not even a consistent estimator. Thus there exists a "gap" between the semiparametric information bounds based on Hellinger differentiability, and the "real" information bounds that consider the amount of smoothness assumed. Birgé and Massart (1995) develop theory for other nonlinear functionals of the form $\nu(P) = \int_0^1 \phi(p(x), p'(x), \ldots, p^{(k)}(x))dx$ for densities p of smoothness s, and for $s \geq 2k + 1/4$ they construct estimators converging at rate $n^{-1/2}$. Moreover they show that $\nu(P)$ cannot be estimated faster than $n^{-\gamma}$ with $\gamma = 4(s-k)/(4s+1)$ when $s < 2k + 1/4$.

These examples and others have been developed further by Robins and Ritov (1997) who make some steps toward development of a "Curse of Dimensionality Appropriate" (or CODA) asymptotic theory of semiparametric models. Robins and Ritov argue via a class of models involving missing data that the existing theory is inadequate and should be altered to incorporate more uniformity in convergence to the limiting distributions. Bickel and Ritov (2000) make a somewhat different suggestion involving regularization of parameters. These ideas deserve further exploration and development.

References

Section 1: Introduction.

1. Andersen, P. K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics* **38**, 67 - 77.
2. Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432 - 452.
3. Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2**, 63 - 74.
4. Beran, R. (1977a). Estimating a distribution function. *Ann. Statist.* **5**, 400 - 404.
5. Beran, R. (1977b). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5**, 445 - 463.
6. Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10**, 647 - 671.
7. Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore. Reprinted (1998), Springer, New York.
8. Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **34**, 187 - 220.
9. Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269 - 276.
10. Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557 - 565.
11. Epstein, B. (1982). Review of *White, J.; Fitting Cox's regression model to survival data using GLIM. Math. Rev.* **82c:62139**.
12. Epstein, B. (1983). Review of *Oakes, D.; Survival times: aspects of partial likelihood. Math. Rev.* **83f:62145**.
13. Finnas, F. and Hoem, J. M. (1980). Starting age and subsequent birth intervals in cohabitational unions in current Danish cohorts, 1975. *Demography* **17**, 275 - 295.
14. Gail, M. H., Santner, T. J., and Brown, C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36**, 255 - 266.
15. Hájek, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Statist.* **33**, 1124 - 1147.
16. Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
17. Kay, R. (1979). Some further asymptotic efficiency calculations for survival data regression models. *Biometrika* **66**, 91 - 96.
18. Louis, T. A., Mosteller, F., and McPeek, B. (1982). Timely topics in statistical methods for clinical trials. *Ann. Rev. Biophys. Bioeng.* **11**, 81 - 104.
19. Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika* **64**, 441 - 448.
20. Oakes, D. (1981). Survival times: aspects of partial likelihood. With a discussion and reply by the author. *Internat. Statist. Rev.* **49**, 235 - 264.
21. Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third*

Berkeley Symp. Math. Statist. Probab. **1**, 187 - 195, University of California Press, Berkeley.

22. Stone, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3**, 267 - 284.
23. Turnbull, B. W. (1981). Review of *Kalbfleisch, J. D. and Prentice, R. L.; The Statistical Analysis of Failure Time Data. Math. Rev.* **81i:62186**.
24. Whitehead, J. (1980). Fitting Cox's regression model to survival data using GLIM. *Appl. Statist.* **29**, 268 - 275.
25. van Eeden, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.* **41**, 172 - 181.

Section 2.1: Missing data models.

26. Breslow, N., McNeney, B., and Wellner, J. A. (2003). Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *Ann. Statist.* **31**, 1110 - 1139.
27. Chen, H. Y. (2002). Double-semiparametric method for missing covariates in Cox regression models. *J. Amer. Statist. Assoc.* **97**, 565 - 576.
28. Chen, H. Y. (2004). Nonparametric and semiparametric models for missing covariates in parametric regression. *J. Amer. Statist. Assoc.* **99**, 1176 - 1189.
29. Nan, B., Emond, M. J., and Wellner, J. A. (2004). Information bounds for Cox regression models with missing data. *Ann. Statist.* **32**, 723 - 753.
30. Robins, J. M., Hsieh, F., and Newey, W. (1995). Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 409 - 424.
31. Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology: Methodological Issues*, 297 - 331. eds. N. P. Jewell, K. Dietz, and V. T. Farewell. Birkhäuser, New York.
32. Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90**, 122 - 129.
33. Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846 - 866.
34. Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90**, 106 - 121.
35. van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge.
36. Wang, Q., Linton, O., and Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.* **99**, 334 - 345.

Section 2.2: Testing and profile likelihood theory.

37. Banerjee, M. (2005). Likelihood ratio tests under local alternatives in regular semiparametric models. *Statist. Sinica* **15**, 635 - 644.

38. Bickel, P. J., Ritov, Y., and Stoker, T. M. (2005a). Tailor-made tests for goodness-of-fit to semiparametric hypotheses. *Ann. Statist.*, to appear.
39. Bickel, P. J., Ritov, Y., and Stoker, T. M. (2005b). Nonparametric testing of an index model. *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock. Cambridge University Press, Cambridge (2005b).
40. Choi, S., Hall, W. J., and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24**, 841 - 861.
41. Härdle, W., Spokoiny, V., and Sperlich, S. (1997). Semiparametric single index versus fixed link function modelling. *Ann. Statist.* **25**, 212 - 243.
42. Horowitz, J. and Härdle, W. (1994). Testing a parametric model against a semiparametric alternative. *Econometric Theory* **10**, 821 - 848.
43. Kauermann, G. and Tutz, G. (2001). Testing generalized linear and semiparametric models against smooth alternatives. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63**, 147 - 166.
44. Lee, B. L., Kosorok, M. R., and Fine, J. P. (2005). The profile sampler. *J. Amer. Statist. Assoc.* **100**, 960 - 969.
45. Murphy, S. A. and van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471 - 1509.
46. Murphy, S. A. and van der Vaart, A. W. (1999). Observed information in semiparametric models. *Bernoulli* **5**, 381 - 412.
47. Murphy, S. A. and van der Vaart, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95**, 449 - 485.
48. Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc* **89**, 501 - 511.
49. Severini, T. A. and Wong, W. H. (1991). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20**, 1768 - 1802.

Section 2.3: Semiparametric mixture model theory.

50. Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Statist.* **27**, 887 - 906.
51. Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1 - 32.
52. Pfanzagl, J. (1988). Consistency of maximum likelihood estimators for certain nonparametric familes, in particular: mixtures. *J. Statist. Plann. Inference* **19**, 137 - 158.
53. Pfanzagl, J. (1990). *Estimation in Semiparametric Models. Lecture Notes in Statistics* **63**. Springer, New York.
54. van der Vaart, A. W. (1988). Estimating a real parameter in a class of semiparametric models. *Ann. Statist.* **16**, 1450 - 1474.
55. van der Vaart, A. W. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24**, 862 - 878.

Section 2.4: Rates of convergence via empirical process methods.

56. Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97**, 113 - 150.
57. Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329 - 375.
58. Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24**, 540 - 568.
59. Shen, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25**, 2555 - 2591.
60. Shen, X. and He, K. (1997). The rate of convergence of PMLEs. *Probab. Math. Statist.* **17**, 21 - 28.
61. Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22**, 580 - 615.
62. van de Geer, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models. *J. Nonparametr. Stat.* **6**, 293 - 310.
63. van der Vaart, A. W. (2002). Semiparametric Statistics. *Lectures on Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999*, *Lecture Notes in Mathematics* **1781**, 330 - 457.
64. van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
65. Wong, W. H. and Severini, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19**, 603 - 632.
66. Wong, W. H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLE's. *Ann. Statist.* **23**, 339 - 362.

Section 2.5: Bayes methods and theory.

67. Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis, 2nd Ed.* Chapman and Hall / CRC Texts, Boca Raton.
68. Freedman, D. (1999). On the Bernstein-Von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27**, 1119 - 1140.
69. Ghosal, S., Ghosh, J. K., and Samanta, T. (1995). On convergence of posterior distributions. *Ann. Statist.* **23**, 2145 - 2152.
70. Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28**, 500 - 531.
71. Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29**, 1233 - 1263.
72. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1999). *Markov Chain MonteCarlo in Practice*. Chapman and Hall, Boca Raton.
73. Huang, T-M. (2004). Convergence rates for posterior distributions and adaptive estimation. *Ann. Statist.* **32**, 1556 - 1593.
74. Kim, Y. and Lee, J. (2004). A Bernstein-von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* **32**, 1492 - 1512.
75. Kleijn, B. J. K. and van der Vaart, A. W. (2002). Misspecification in Infinite -Dimensional Bayesian statistics. *Technical Report* **2002-16**, Free University, Amsterdam.
76. Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika* **90**, 319 -

326.
77. Lenk, P. J. (1999). Bayesian inference for semiparametric regression using a Fourier representation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 863 - 879.
78. Müller, P. and Roeder, K. A. (1997). Bayesian semiparametric model for case-control studies with errors in variables. *Biometrika* **84**, 523 - 537.
79. Newton, M. A., Czado, C., and Chappell, R. (1996). Bayesian inference for semiparametric binary regression. *J. Amer. Statist. Assoc.* **91**, 142 - 153.
80. Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods* Springer, New York.
81. Shen, X. (2002). Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97**, 222 - 235.
82. Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29**, 687 - 714.
83. Sinha, D., Ibrahim, J. G., and Chen, M-H. (2003). A Bayesian justification of Cox's partial likelihood. *Biometrika* **90**, 629 - 641.
84. Vidakovic, B. (1998). Wavelet-based nonparametric Bayes methods. *Practical Nonparametric and Semiparametric Bayesian Statistics*, 133 - 155, *Lecture Notes in Statistics* **133**, Springer, New York.

Section 2.6: Model selection methods.

85. Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113**, 301 - 413.
86. Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. *Festschrift for Lucien Le Cam*, 55 - 87, Springer, New York.
87. Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329 - 375.
88. Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* **3**, 203 - 268.
89. Bunea, F. (2004). Consistent covariate selection and post model selection inference in semiparametric regression. *Ann. Statist.* **32**, 898 - 927.
90. Bunea, F. and McKeague, I. W. (2005). Covariate selection for semiparametric hazard function regression models. *J. Multivariate Anal.* **92**, 186 - 204.
91. Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Ann. Statist.* **30**, 74 - 99.
92. Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27**, 1536 - 1563.
93. Massart, P. (2000). Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse Math.* **(6) 9**, 245 - 303.
94. Raftery, A. E., Madigan, D., and Volinsky, C. T. (1996). Accounting for model uncertainty in survival analysis improves predictive performance. *Bayesian Statistics 5 – Proceedings of the Fifth Valencia International Meeting, Valencia, Spain.* Oxford Science Publications, Oxford.
95. Raftery, A. E., Madigan, D., Volinsky, C. T., and Kronmal, R. A. (1997). Bayesian model averaging in proportional hazards models: predicting the

risk of a stroke. *Appl. Statist.* **46**, 443 - 448.
96. Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statist. in Medicine* **16**, 385 - 395.

Section 2.7: Empirical likelihood.

97. Koshevnik, Yu. A. and Levit, B. Ya. (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Appl.* **21**, 738 - 753.
98. Lazar, N. A. and Mykland, P. A. (1999). Empirical likelihood in the presence of nuisance parameters. *Biometrika* **86**, 203 - 211.
99. Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237 - 249.
100. Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90 - 120.
101. Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725 - 1747.
102. Qin, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21**, 1182 - 1196.
103. Qin, J. (1998). Semiparametric likelihood based method for goodness of fit tests and estimation in upgraded mixture models. *Scand. J. Statist.* **25**, 681 - 691.
104. Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300 - 325.
105. Qin, J. and Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *Canad. J. Statist.* **23**, 145 - 159.
106. Zou, F., Fine, J. P., and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika* **89**, 61 - 75.

Section 2.8: Transformation and frailty models.

107. Bickel, P. J. (1986). Efficient testing in a class of transformation models. *Proceedings of the 45th Session of the International Statistical Institute* 23.3.63 - 23.3.81. ISI, Amsterdam.
108. Bickel, P. J. and Ritov, Y. (1997). Local asymptotic normality of ranks and covariates in transformation models. *Festschrift for Lucien Le Cam*, 43 - 54, Springer, New York.
109. Cai, T. X. and Moskowitz, C. S. (2004). Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics* **5**, 573 - 586.
110. Chen, X., Fan, Y., and Tsyrennikov, V. (2004). Efficient estimation of semiparametric multivariate copula models. *Working Paper Nol. 04-W20*, Department of Economics, Vanderbilt University. Available at: *http://www.vanderbilt.edu/Econ/wparchive/working04.html*
111. Clayton, D. G. and Cuzick, J. (1985). The semi-parametric Pareto model for regression analysis of survival times. *Bull. Int. Statist. Inst.* **51**, 23.3.175 - 23.3.180.
112. Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34**, 187 - 220.

113. Horowitz, J. L. (1998). Semiparametric Methods in Econometrics. *Lecture Notes in Statistics* **131**, Springer-Verlag, New York.
114. Hsieh, F. and Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.* **24**, 25 - 40.
115. Klaassen, C. A. J. (2005). A Sturm-Liouville problem in semiparametric transformation models. Festschrift volume for Kjell Doksum, to appear.
116. Klaassen, C. A. J. and Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favorable. *Bernoulli* **3**, 55 - 77.
117. Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics* **58**, 287 - 297.
118. Mallick, B. K. and Walker, S. (2003). A Bayesian semiparametric transformation model incorporating frailties. *Model selection, model diagnostics, empirical Bayes and hierarchial Bayes, Special Issue II, J. Statist. Plann. Inference* **112**, 159 - 174.
119. Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182 - 198.
120. Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183 - 214.
121. Zou, K. H. and Hall, W. J. (2000). Two transformation models for estimating an ROC curve derived from continuous data. *J. Appl. Stat.* **27**, 621 - 631.
122. Zou, K. H. and Hall, W. J. (2002). Semiparametric and parametric transformation models for comparing diagnostic markers with paired design. *J. Appl. Stat.* **29**, 803 - 816.

Section 2.9: Semiparametric regression models.

123. Cuzick, J. (1992). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Ann. Statist.* **20**, 1129 - 1136.
124. Fan, J., Peng, H., and Huang, T. (2005). Semilinear high-dimensional model for normalization of microarray data: a theoretical analysis and partial consistency. With discussion. *J. Amer. Statist. Assoc.* **100**, 781 - 813.
125. Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27**, 1536 - 1563.
126. Huang, J., Wang, D., and Zhang, C-H. (2005). A two-way semilinear model for normalization and analysis of cDNA microarray data. *J. Amer. Statist. Assoc.* **100**, 814 - 829.
127. Nielsen, J. P., Linton, O., and Bickel, P. J. (1998). On a semiparametric survival model with flexible covariate effect. *Ann. Statist.* **26**, 215 - 241.
128. Ritov, Y. and Bickel, P. J. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist.* **18**, 925 - 938.
129. Sasieni, P. (1992a). Information bounds for the conditional hazard ratio in a nested family of regression models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **54**, 617 - 635.
130. Sasieni, P. (1992b). Non-orthogonal projections and their application to calculating the information in a partly linear Cox model. *Scand. J. Statist.*

19, 215 - 233.
131. Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21**, 1486 - 1521. Correction and addendum. *Ann. Statist.* **23**, 1862 - 1863.
132. Seifu, Y., Severini, T. A., and Tanner, M. A. (1999). Semiparametric Bayesian inference for regression models. *Canad. J. Statist.* **27**, 719 - 734.

Section 2.10: Extensions to non-i.i.d. data.

133. Bickel, P. J. (1993). Estimation in semiparametric models. *Multivariate Analysis: Future Directions*, 55 - 73.
134. Bickel, P. J. and Kwon, J. (2001). Inference for semiparametric models: some questions and an answer. With comments and a rejoinder by the authors. *Statist. Sinica* **11**, 863 - 960.
135. Drost, F. C. and Klaassen, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. *J. Econometrics* **81**, 193 - 221.
136. Drost, F. C., Klaassen, C. A. J., and Werker, B. J. M. (1997). Adaptive estimation in time-series models. *Ann. Statist.* **25**, 786 - 817.
137. Gassiat, E. (1990). Estimation semi-paramétrique d'un modèle autorégressif stationnaire multiindice non nécessairement causal. [Semiparametric estimation of a not necessarily causal multi-index stationary autoregressive model]. *Ann. Inst. H. Poincaré Probab. Statist.* **26**, 181 - 205.
138. Gassiat, E. (1993). Adaptive estimation in noncausal stationary AR processes. *Ann. Statist.* **21**, 2022 - 2042.
139. Greenwood, P. E., Müller, U. U., and Wefelmeyer, W. (2004). An introduction to efficient estimation for semiparametric time series. *Parametric and semiparametric models with applications to reliability, survival analysis, and quality of life*, 253 - 269, Stat. Ind. Technol., Birkhäuser, Boston.
140. Greenwood, P. E. and Wefelmeyer, W. (1995). Efficiency of empirical estimators for Markov chains. *Ann. Statist.* **23**, 132 - 143.
141. Kessler, M., Schick, A., and Wefelmeyer, W. (2001). The information in the marginal law of a Markov chain. *Bernoulli* **7**, 243 - 266.
142. Koul, H. L. and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Ann. Statist.* **24**, 1025 - 1052.
143. Koul, H. L. and Schick, A. (1997). Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* **3**, 247 - 277.
144. Kreiss, J. P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15**, 112 - 133.
145. McNeney, B. and Wellner, J. A. (2000). Application of convolution theorems in semiparametric models with non-i.i.d. data. *Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998). J. Statist. Plann. Inference* **91**, 441 - 480.
146. Murphy, S. A. and van der Vaart, A. W. (1996). Likelihood inference in the errors-in-variables model. *J. Multivariate Anal.* **59**, 81 - 108.
147. Schick, A. and Wefelmeyer, W. (1999). Efficient estimation of invariant distributions of some semiparametric Markov chain models. *Math. Methods Statist.* **8**, 426 - 440.
148. Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third*

Berkeley Symp. Math. Statist. Prob. **1**, 187 - 195. University of California Press, Berkeley.

Section 2.11: Critiques and possible alternative theories.

149. Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* **50**, 381 - 393.
150. Bickel, P. J. and Ritov, Y. (2000). Non- and semiparametric statistics: compared and contrasted. *J. Statist. Plann. Inference* **91**, 209 - 228.
151. Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23**, 11 - 29.
152. Ritov, Y. and Bickel, P. J. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist.* **18**, 925 - 938.
153. Robins, J.M. and Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16**, 285 - 319.