STANFORD UNIVERSITY

DEPARTMENT OF STATISTICS

DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, February 8, 2000
Sequoia Hall Rm. 200
(Cookies at 3:45 in 1st Floor Lounge)

Florentina Bunea
Dept. of Statistics
University of Washington

A Model Selection Approach to Partially Linear Regression

We extend the model selection approach proposed by Barron, Birgé and Massart (1999) for nonparametric regression to partially linear regression. That is, we consider the model $Y = X'\beta + f(T) + W$, where X belongs to R^q, T belongs to R, W is the error independent of (X, T) and f is a function of unknown smoothness. This model has received considerable attention in the literature, and it is mostly used in cases in which the parameter of interest is the linear component, whereas the variables appearing in the nonlinear part are viewed as confounders, hence f is regarded as a nuisance parameter. We study this model in two different cases.

Case A. The number of covariates appearing in the linear part is a priori given, say q.

Case B. We have available q possible regressors for the linear part, but only an unknown (possibly much smaller) subset of them are relevant for Y, hence we would like to *select* it.

We propose a penalized least squares approach and, in both cases, we obtain finite sample upper bounds for the risk of the estimator and, as a consequence, the consistency of the estimators of β and f at the optimal nonparametric rate. In **Case B**, for $f \equiv 0$, we derive rates of convergence for estimators of β corresponding to BIC and AIC. We also discuss the distributional properties of $\hat{\beta}$ in Case A.