IEOR 290A – Lecture 11
Semiparametric Models

1 Nuisance Parameters

Consider the basic linear model $y_i = x_i'\beta + \epsilon_i$, where ϵ_i is i.i.d. noise with zero noise with finite variance. So far, we have focused on the question of estimating β; but, we could also ask the question whether it is possible to something about ϵ_i. The reason that we have not addressed this issue is that, because the ϵ_i in this model represent random noise with zero mean, we do not gain any information for the purposes of model prediction (i.e., estimating $\mathbb{E}[y_i|x_i] = x_i'\beta$) by estimating the ϵ_i (or alternatively information about its distribution). However, if we are interested in understanding the uncertainty of our model predictions, then it is valuable to estimate the distribution of ϵ_i.

These ϵ_i are examples of *nuisance parameters*, which are any parameters that are not directly of interest but must be considered in the estimation. (Note that the designation of a parameter as a nuisance parameter is situationally dependent – in some applications, the nuisance parameter is also of interest.) In general, we can have situations in which there are a finite number of nuisance parameters or even an infinite number of nuisance parameters. There is no standard approach to handling nuisance parameters in regression problems. One approach is to estimate the nuisance parameters anyways, but unfortunately it is not always possible to estimate the nuisance parameters. Another approach is to consider the nuisance parameters as "worst-case disturbances" and use minmax estimators, which can be thought of as game-theoretic M-estimators.

1.1 Gaussian Noise in Linear Model

Consider the linear model in the situation where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ for some unknown variance σ^2. Recall that the M-estimator was given by

$$\hat{\beta} = \arg\max_{\beta} \sum_{i=1}^{n} \Big(-(y_i - x_i'\beta)^2/(2\sigma^2) - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log(2\pi)\Big).$$

In this case, the nuisance parameter is σ^2. The way this parameter was handled was to observe that the maximizer is independent of σ^2, which allowed us to rewrite the M-estimator as

$$\hat{\beta} = \arg\max_{\beta} \sum_{i=1}^{n} -(y_i - x_i'\beta)^2 = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - x_i'\beta)^2 = \arg\min_{\beta} \|Y - X\beta\|_2^2.$$

1.2 Generic Noise in Linear Model

Now consider the linear model in the case where ϵ_i is a generic zero mean distribution, meaning that it is of some unknown distribution. It turns out that we can estimate each ϵ_i in a consistent manner. Suppose that we assume

1. the norm of the x_i is deterministically bounded: $\|x_i\| \leq M$ for a finite $M < \infty$;

2. conditions under which OLS $\hat{\beta}$ is a consistent estimate of β.

Then we can use OLS to estimate the ϵ_i. Define

$$\hat{\epsilon}_i = y_i - x_i'\hat{\beta},$$

and note that

$$\begin{aligned}
|\hat{\epsilon}_i - \epsilon_i| &= |(y_i - x_i'\hat{\beta}) - \epsilon_i| \\
&= |x_i'\beta + \epsilon_i - x_i'\hat{\beta} - \epsilon_i| \\
&= |x_i'(\beta - \hat{\beta})| \\
&\leq \|x_i\| \cdot \|(\beta - \hat{\beta})\|,
\end{aligned}$$

where in the last line we have used the Cauchy-Schwarz inequality. And because of our assumptions, we have that $|\hat{\epsilon}_i - \epsilon_i| = O_p(1/\sqrt{n})$.

Now in turn, our estimates of ϵ_i can be used to estimate other items of interest. For example, we can use our estimates of $\hat{\epsilon}_i$ to estimate population parameters such as variance:

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2.$$

This estimator is consistent:

$$\begin{aligned}
|\hat{\sigma^2} - \sigma^2| &= \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 - \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 + \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 - \sigma^2 \right| \\
&\leq \left| \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 - \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 \right| + \left| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 - \sigma^2 \right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \left| \hat{\epsilon}_i^2 - \epsilon_i^2 \right| + \left| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 - \sigma^2 \right|.
\end{aligned}$$

where we have made use of the triangle inequality in the second and third lines. Next note that $|\hat{\epsilon}_i^2 - \epsilon_i^2| = O_p(1/\sqrt{n})$ by a version of the continuous mapping theorem and that $|\frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2 - \sigma^2| = O_p(1/\sqrt{n})$ because of the CLT. Thus, we have that $|\hat{\sigma^2} - \sigma^2| = O_p(1/\sqrt{n})$.

2 Partially Linear Model

Consider the following model

$$y_i = x_i'\beta + g(z_i) + \epsilon_i,$$

where $y_i \in \mathbb{R}$, $x_i, \beta \in \mathbb{R}^p$, $z_i \in \mathbb{R}^q$, $g(\cdot)$ is an unknown nonlinear function, and ϵ_i are noise. The data x_i, z_i are i.i.d., and the noise has conditionally zero mean $\mathbb{E}[\epsilon_i|x_i, z_i] = 0$ with unknown and bounded conditional variance $\mathbb{E}[\epsilon_i^2|x_i, z_i] = \sigma^2(x_i, z_i)$. This is known as a partially linear model because it consists of a (parametric) linear part $x_i'\beta$ and a nonparametric part $g(z_i)$. One can think of the $g(\cdot)$ as an infinite-dimensional nuisance parameter.

3 Single-Index Model

Consider the following model

$$y_i = g(x_i'\beta) + \epsilon_i,$$

where $y_i \in \mathbb{R}$, $x_i, \beta \in \mathbb{R}^p$, $g(\cdot)$ is an unknown nonlinear function, and ϵ_i are noise. The data x_i are i.i.d., and the noise has conditionally zero mean $\mathbb{E}[\epsilon_i|x_i] = 0$. Such single-index models can be used for asset pricing, and here the $g(\cdot)$ can be thought of as an infinite-dimensional nuisance parameter.