Lecture Note 5: LAD

Economics 231a Fall, 2005

Motivation

Suppose that u has a continuous distribution and has median equal to m. It can then be shown that

$$\underset{t}{\operatorname{argmin}}\, E\left[|u - t|\right] = m$$

Now suppose that we have a linear regression model

$$y_i = x_i'\beta + u_i$$

with the restriction that the conditional median of u_i given x_i is equal to zero. Because

$$E\left[|y_i - t(x_i)|\,|\, x_i\right]$$

is minimized at $t(x_i) = x_i'\beta$, we can see that

$$\underset{b}{\operatorname{argmin}}\, E\left[|y_i - x_i'b|\right] = \beta$$

Sample analog is then given by

$$\widehat{\beta} = \underset{b}{\operatorname{argmin}}\, \frac{1}{n}\sum_{i=1}^{n} |y_i - x_i'b|$$

Asymptotic Distribution

Despite the nondifferentiability problem, our asymptotic machinery for analysis of m-estimator turns out still valid. Note that the derivative of $|y_i - x_i'b|$, when it exists, is equal to

$$1\left(y_i - x_i'b < 0\right)x_i - 1\left(y_i - x_i'b > 0\right)x_i$$

Therefore the LAD can be roughly understood as a method of moments estimator corresponding to the moment restriction

$$s\left(w_i, b\right) = 1\left(y_i - x_i'b < 0\right)x_i - 1\left(y_i - x_i'b > 0\right)x_i$$

According to our machinery, we should have

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} N\left(0, A^{-1}BA'^{-1}\right)$$

where

$$\begin{aligned} A &= E\left[\frac{\partial s(w_i,\beta)}{\partial b'}\right] \\ B &= E\left[s(w_i,\beta)\, s(w_i,\beta)'\right] \end{aligned}$$

We have

$$\begin{aligned} B &= E\left[s(w_i,\beta)\, s(w_i,\beta)'\right] = E\left[\left(1(u_i<0)-1(u_i>0)\right)^2 x_i x_i'\right] \\ &= E\left[x_i x_i'\right] \end{aligned} \tag{1}$$

and

$$A = E\left[\frac{\partial s(w_i,\beta)}{\partial b'}\right] = E\left[\frac{\partial\left(1(y_i - x_i'b<0)-1(y_i-x_i'b>0)\right)}{\partial b'}\right] x_i$$

But $1(y_i - x_i'b<0)-1(y_i-x_i'b>0)$ is not differentiable, so our machinery is not directly applicable. We cheat, and use an alternative definition of A:

$$A = \frac{\partial}{\partial b'}E\left[s(w_i,b)\right]\Big|_{b=\beta} = \frac{\partial}{\partial b'}E\left[\left(1(y_i-x_i'b<0)-1(y_i-x_i'b>0)\right)x_i\right]\Big|_{b=\beta}$$

Note that

$$\begin{aligned} & E\left[\left(1(y_i-x_i'b<0)-1(y_i-x_i'b>0)\right)\cdot x_i \mid x_i\right] \\ = \; & E\left[\left(1(u_i - x_i'(b-\beta)<0)-1(u_i-x_i'(b-\beta)>0)\right)\mid x_i\right]\cdot x_i \\ = \; & \left(F(x_i'(b-\beta)\mid x_i) - (1-F(x_i'(b-\beta)\mid x_i))\right)\cdot x_i \\ = \; & \left(2F(x_i'(b-\beta)\mid x_i)-1\right)\cdot x_i \end{aligned}$$

where $F(\cdot\mid x_i)$ denotes the conditional CDF of u_i given x_i. We therefore have

$$E\left[\left(1(y_i-x_i'b<0)-1(y_i-x_i'b>0)\right)x_i\right] = E\left[\left(2F(x_i'(b-\beta)\mid x_i)-1\right)\cdot x_i\right]$$

and

$$\begin{aligned} \frac{\partial}{\partial b'}E\left[\left(1(y_i-x_i'b<0)-1(y_i-x_i'b>0)\right)x_i\right] &= \frac{\partial}{\partial b'}E\left[\left(2F(x_i'(b-\beta)\mid x_i)-1\right)\cdot x_i\right] \\ &= E\left[\frac{\partial}{\partial b'}\left(2F(x_i'(b-\beta)\mid x_i)-1\right)\cdot x_i\right] \\ &= E\left[2f(x_i'(b-\beta)\mid x_i)\,x_i'\cdot x_i\right] \\ &= 2E\left[f(x_i'(b-\beta)\mid x_i)\cdot x_i'x_i\right] \end{aligned}$$

where $F(\cdot\mid x_i)$ denotes the conditional PDF of u_i given x_i. Therefore,

$$\begin{aligned} A &= \frac{\partial}{\partial b'}E\left[s(w_i,b)\right]\Big|_{b=\beta} \\ &= 2\,E\left[f(x_i'(b-\beta)\mid x_i)\cdot x_i'x_i\right]\big|_{b=\beta} \\ &= 2E\left[f(x_i'(\beta-\beta)\mid x_i)\cdot x_i'x_i\right] \\ &= 2E\left[f(0\mid x_i)\cdot x_i'x_i\right] \end{aligned} \tag{2}$$

Combining (1) and (2), we obtain

$$\sqrt{n}\left(\widehat{\beta}-\beta\right) \overset{d}{\to} N\left(0, \frac{1}{4}\left(E\left[f\left(0|\, x_i\right)\cdot x_i' x_i\right]\right)^{-1} E\left[x_i' x_i\right]\left(E\left[f\left(0|\, x_i\right)\cdot x_i' x_i\right]\right)^{-1}\right)$$

If u_i and x_i are independent of each other, then we have $f\left(0|\, x_i\right) = f\left(0\right)$, and

$$\sqrt{n}\left(\widehat{\beta}-\beta\right) \overset{d}{\to} N\left(0, \frac{1}{4f\left(0\right)^2}\left(E\left[x_i' x_i\right]\right)^{-1}\right)$$

Estimation of asymptotic variance requires knowledge of or estimation of $f\left(0\right)$. This object may be estimated based on residuals $\{\widehat{u}_1, \ldots, \widehat{u}_n\}$. Below, we discuss how it could be estimated if $\{u_1, \ldots, u_n\}$ were observed.

Digression 1: Nonparametric Density Estimation

Suppose that you are given a random sample $w_1, \ldots, w_n$ from a common PDF $f\left(\cdot\right)$. Suppose that you are interested in estimating $f\left(c\right)$ for some $c \in \mathbb{R}$ without any parametric knowledge. For this purpose, note that

$$f\left(c\right) = \lim_{h\to\infty} \frac{F\left(c+h\right)-F\left(c-h\right)}{2h} \approx \frac{F\left(c+h\right)-F\left(c-h\right)}{2h} = \frac{\Pr\left[c-h < w_i < c+h\right]}{2h}$$

Therefore, a reasonable estimate of $f\left(c\right)$ seems to be the sample analog

$$\widehat{f}\left(c\right) = \frac{1}{2h}\frac{1}{n}\sum_{i=1}^{n} 1\left\{c-h < w_i < c+h\right\} = \frac{1}{2nh}\sum_{i=1}^{n} 1\left\{c-h < w_i < c+h\right\}$$

Letting

$$K\left(u\right) = \frac{1}{2} 1\left\{-1 < u < 1\right\}$$

we can rewrite

$$\widehat{f}\left(c\right) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{w_i - c}{h}\right)$$

Note that

$$\begin{aligned} \int K\left(u\right) du &= 1 \\ \int uK\left(u\right) du &= 0 \end{aligned}$$

and hence, can be understood as an auxiliary density function, or kernel, for estimation of $f\left(\cdot\right)$. This suggests that we can use some other kernels such as

$$\begin{aligned} K\left(u\right) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \\ K\left(u\right) &= \frac{3}{4\sqrt{5}}\left(1-\frac{u^2}{5}\right)\cdot 1\left\{\left|u\right| < \sqrt{5}\right\} \end{aligned}$$

Note that

$$\begin{aligned}
E\left[\widehat{f}(c)\right] &= E\left[\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{w_i-c}{h}\right)\right] = \frac{1}{h}E\left[K\left(\frac{w_i-c}{h}\right)\right] = \frac{1}{h}\int K\left(\frac{u-c}{h}\right)f(u)\,du \\
&= \int K(z)\,f(c+hz)\,dz \approx \int K(z)\left(f(c)+f'(c)\,hz+\frac{1}{2}f''(c)\,(hz)^2\right)dz \\
&= f(c)\int K(z)\,dz + hf'(c)\int zK(z)\,dz + \frac{1}{2}h^2 f''(c)\int z^2K(z)\,dz \\
&= f(c)+\frac{\kappa_2 f''(c)}{2}h^2
\end{aligned}$$

for

$$\kappa_2 = \int u^2 K(u)\,du$$

In order to have a small bias, we would like to have h “small”.

Also note that

$$\text{Var}\left(\widehat{f}(c)\right) = \text{Var}\left(\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{w_i-c}{h}\right)\right) = \frac{1}{nh^2}\,\text{Var}\left(K\left(\frac{w_i-c}{h}\right)\right)$$

Because

$$\begin{aligned}
E\left[K\left(\frac{w_i-c}{h}\right)^2\right] &= \int K\left(\frac{u-c}{h}\right)^2 f(u)\,du \\
&= h\int K(z)^2 f(c+hz)\,dz \\
&\approx h\int K(z)^2\left(f(c)+f'(c)\,hz+\frac{1}{2}f''(c)\,(hz)^2\right)dz \\
&= hf(c)\int K(z)^2\,dz + h^2f'(c)\int zK(z)^2\,dz + \frac{1}{2}h^3f''(c)\int z^2K(z)^2\,dzdz
\end{aligned}$$

we have

$$\begin{aligned}
\text{Var}\left(K\left(\frac{w_i-c}{h}\right)\right) &= hf(c)\int K(z)^2\,dz + h^2f'(c)\int zK(z)^2\,dz + \frac{1}{2}h^3f''(c)\int z^2K(z)^2\,dzdz \\
&\quad -\left(h\cdot\left(f(c)+\frac{\kappa_2 f''(c)}{2}h^2\right)\right)^2 \\
&= hf(c)\int K(z)^2\,dz + o(h)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{Var}\left(\widehat{f}(c)\right) &= \frac{1}{nh^2}\,\text{Var}\left(K\left(\frac{w_i-c}{h}\right)\right) \\
&= \frac{1}{nh^2}\left\{hf(c)\int K(z)^2\,dz + o(h)\right\} \\
&\approx \frac{f(c)\int K(z)^2\,dz}{nh}
\end{aligned}$$

and we would like to have nh "large".

Note that the mean squared error is approximately equal to

$$\left(\frac{\kappa_2 f''(c)}{2}h^2\right)^2 + \frac{f(c)\int K(z)^2\,dz}{nh}$$

We would like to choose $h = O\left(n^{-\delta}\right)$ such that the MSE is minimized. Note that the bias squared is of order $O\left(\left(h^2\right)^2\right) = O\left(n^{-4\delta}\right)$, and the variance is of order $O\left(n^{-1+\delta}\right)$. It follows that

$$\text{MSE} = O\left(n^{-4\delta}\right) + O\left(n^{-1+\delta}\right) = O\left(\max\left(n^{-4\delta}, n^{-1+\delta}\right)\right)$$

MSE is minimized when $n^{-4\delta}$ and $n^{-1+\delta}$ are equalized, i.e., $\delta = \frac{1}{5}$. This leads to the usual suggestion in the literature: $h = O\left(n^{-1/5}\right)$

Digression 2: Multivariate Density Estimation

If $\dim(w_i) = K$, we have

$$f(c_1, \ldots, c_K) = \lim_{h\to\infty} \frac{\Pr\left[c_1 - h < w_{i,1} < c_1 + h, \ldots, c_K - h < w_{i,K} < c_K + h\right]}{(2h)^K}$$

which suggests an estimator of the form

$$\widehat{f}(c) = \frac{1}{n(2h)^K}\sum_{i=1}^{n} 1\left\{c_1 - h < w_{i,1} < c_1 + h, \ldots, c_K - h < w_{i,K} < c_K + h\right\}$$

Letting

$$\begin{aligned} K(u_1, \ldots, u_K) &= \frac{1}{2^K} 1\{-1 < u_1 < 1, \ldots, -1 < u_K < 1\} \\ &= \prod_{k=1}^{K}\left(\frac{1}{2}1\{-1 < u_k < 1\}\right) = \prod_{k=1}^{K}\mathcal{K}(u_k) \end{aligned}$$

we can rewrite

$$\widehat{f}(c) = \frac{1}{nh^K}\sum_{i=1}^{n} K\left(\frac{w_i - c}{h}\right)\left(\qquad = \frac{1}{nh^K}\sum_{i=1}^{n}\left(\prod_{k=1}^{K}\mathcal{K}\left(\frac{w_{i,k} - c_k}{h}\right)\right)\right)$$

This suggests that we can use some other kernel as in the univariate case as well.

Digression 3: Nonparametric Regression

Note that

$$\begin{aligned} E\left[y_i \mid x_i = c\right] &= \lim_{h\to 0} E\left[y_i \mid c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right] \\ &= \lim_{h\to 0} \frac{E\left[y_i \cdot 1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\}\right]}{E\left[1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\}\right]} \end{aligned}$$

Therefore, we have

$$E\left[y_i \middle| x_i = c\right] \approx \frac{E\left[y_i \cdot 1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\}\right]}{E\left[1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\}\right]}$$

for some small h. Because

$$\begin{aligned} & E\left[y_i \cdot 1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\}\right] \\ \approx\ & \frac{1}{n}\sum_{i=1}^{n} y_i \cdot 1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\} \\ =\ & \frac{1}{n}\sum_{i=1}^{n} y_i \cdot K\left(\frac{x_i - c}{h}\right) \end{aligned}$$

and

$$\begin{aligned} & E\left[1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\}\right] \\ \approx\ & \frac{1}{n}\sum_{i=1}^{n} 1\left\{c_1 - h < x_{i,1} < c_1 + h, \ldots, c_K - h < x_{i,K} < c_K + h\right\} \\ =\ & \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x_i - c}{h}\right) \end{aligned}$$

we obtain the kernel regression estimator

$$\frac{\frac{1}{n}\sum_{i=1}^{n} y_i \cdot K\left(\frac{x_i - c}{h}\right)}{\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x_i - c}{h}\right)}$$