

Non-parametric Regression

- The outcome is a function of the regressor.
- Economic theory usually does not predict functional forms, but often implies constraints.

$$Y_i = f(X_i) + \varepsilon_i$$

- Parametric estimation involves replacing $f(x)$ with a polynomial and estimating the model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_q X_i^q + \varepsilon_i$$

$$\text{where } q \geq 3$$

- Non-parametric regression generally involves using some type of smoothing to estimate $f(x)$. The simplest smoothing would be a moving average, where $f(x^*)$ is approximated by the average of points around x^* .
- A popular non-parametric regression is the Gaussian kernel regression, which essentially is a weighted average of points close to x .

$$\hat{f}(x) = \sum_{i=1}^n w(x, x_i) y_i, \quad w(x, x_i) = \frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{j=1}^n K\left(\frac{x_j - x}{h}\right)}$$

$$\text{where } K(a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}} \quad \& \quad h = \text{bandwidth } h$$

- K is a kernel, there are many of them, but Gaussian is popular. They integrate to one and are symmetric.
- The bandwidth h is a smoothing parameter. The bigger the h , the more points will get significant weight, this will lead to under fitting. Small h will imply the opposite and lead to overfitting.
- Kernel regression does not require any estimation of parameters (the only parameter is $h = \text{bandwidth}$). Compute $\hat{f}(x)$ on a finite grid on a relevant support.
- Another non-parametric regression is the locally weighted linear regression, where we are fitting many linear regression to approximate $f(x)$.
- Local weighted regression motivated by Taylor's theorem as it suggests that $f(x)$ at a point can be approximated by a polynomial.

$$Y_i = \beta_0(x) + \beta_1(x) X_i + \varepsilon_i \quad \text{for } x \in [x_i - \delta, x_i + \delta]$$

$$\hookrightarrow \beta(x) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \left\{ \sum_{i=1}^n w(x, x_i) (y_i - \beta_0 - \beta_1 X_i)^2 \right\}$$

where $w(x, x_i) = e^{-(x_i - x)^2 / 2\tau^2}$

- Need to estimate parameters each time you need to do an prediction. Would need to estimate many parameters before plotting estimated $f(x)$. Closer points gets more weight.
- As seen by the locally-weighted regression example, "non-parametric regression" does indeed have parameters. However now the parameters are not fixed, but can rather increase in the sample size.