

Frontier Topics in Empirical Economics: Week 2

Non-parametric Method

Zibin Huang¹

¹College of Business, Shanghai University of Finance and Economics

November 30, 2023

Non-parametric Method: Introduction

- Common Parametric Models
 - Linear Model: $y = X\beta + e, e \sim N(0, \sigma^2)$;
 - Probit/Logit Model: $P(y=1|X) = G(X\beta)$ where G is a nonlinear function
- Explicit Parametric Structure for Distribution
- Common Estimator
 - OLS, MLE, Nonlinear LS, Efficient GMM etc.
- Key Properties of the Estimator
 - Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models

Linear Model: $y = X\beta + e, e \sim N(0, \sigma^2)$;

Probit/Logit Model: $P(y=1|X) = G(X\beta)$ where G is a nonlinear function

- Explicit Parametric Structure for Distribution

- Common Estimator

OLS, MLE, Nonlinear LS, Efficient GMM etc.

- Key Properties of the Estimator

Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models

Linear Model: $y = X\beta + e, e \sim N(0, \sigma^2)$;

Probit/Logit Model: $P(y=1|X) = G(X\beta)$ where G is a nonlinear function

- Explicit Parametric Structure for Distribution

- Common Estimator

OLS, MLE, Nonlinear LS, Efficient GMM etc.

- Key Properties of the Estimator

Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models

Linear Model: $y = X\beta + e, e \sim N(0, \sigma^2)$;

Probit/Logit Model: $P(y=1|X) = G(X\beta)$ where G is a nonlinear function

- Explicit Parametric Structure for Distribution

- Common Estimator

OLS, MLE, Nonlinear LS, Efficient GMM etc.

- Key Properties of the Estimator

Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models
 - Linear Model: $y = X\beta + e, e \sim N(0, \sigma^2)$;
 - Probit/Logit Model: $P(y=1|X) = G(X\beta)$ where G is a nonlinear function
- Explicit Parametric Structure for Distribution
- Common Estimator
 - OLS, MLE, Nonlinear LS, Efficient GMM etc.
- Key Properties of the Estimator
 - Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- In linear regression, we have to assume that CEF is linear
- Why linear? Simple? Why not $y = x^3 \ln x + e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression

Non-parametric Method: Introduction

- In linear regression, we have to assume that CEF is linear
- Why linear? Simple? Why not $y = x^3 \ln x + e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression

Non-parametric Method: Introduction

- In linear regression, we have to assume that CEF is linear
- Why linear? Simple? Why not $y = x^3 \ln x + e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression

Non-parametric Method: Introduction

- In linear regression, we have to assume that CEF is linear
- Why linear? Simple? Why not $y = x^3 \ln x + e$?
- What if linear specification is wrong?
 - Everything collapses. No data can save.
 - It becomes only a linear approximation
 - For example, if true model is Logit, but not linear regression

Non-parametric Method: Introduction

- In linear regression, we have to assume that CEF is linear
- Why linear? Simple? Why not $y = x^3 \ln x + e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression

Non-parametric Method: Introduction

- In linear regression, we have to assume that CEF is linear
- Why linear? Simple? Why not $y = x^3 \ln x + e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression

Non-parametric Method: Introduction

- In linear regression, we have to assume that CEF is linear
- Why linear? Simple? Why not $y = x^3 \ln x + e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression

Non-parametric Method: Introduction

- Potential Outcome Model is intrinsically NON-parametric!!!
- There are many other statistical modeling methods
- Non-parametric, semi-parametric
- To understand tools beyond linear regression!

Non-parametric Method: Introduction

- Potential Outcome Model is intrinsically NON-parametric!!!
- There are many other statistical modeling methods
- Non-parametric, semi-parametric
- To understand tools beyond linear regression!

Non-parametric Method: Introduction

- Potential Outcome Model is intrinsically NON-parametric!!!
- There are many other statistical modeling methods
 - Non-parametric, semi-parametric
 - To understand tools beyond linear regression!

Non-parametric Method: Introduction

- Potential Outcome Model is intrinsically NON-parametric!!!
- There are many other statistical modeling methods
- Non-parametric, semi-parametric
- To understand tools beyond linear regression!

Non-parametric Method: Introduction

- Potential Outcome Model is intrinsically NON-parametric!!!
- There are many other statistical modeling methods
- Non-parametric, semi-parametric
- To understand tools beyond linear regression!

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E(y_i|x_i)$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E(y_i | x_i)$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E(y_i|x_i)$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E(y_i|x_i)$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E y_i | x_i$ **without imposing any functional form assumption**

Non-parametric Method: Introduction

- Notation: $x_i; y_i$ denotes random variable; $X_i; Y_i$ denotes realizations; $x; y$ denotes random variables or some value of the random variables
 - Realizations are given (sample), they are NOT random in our context
- $$\int_{D_x} x < \int_{D_x} X_i dx < \int_{D_x} X_i D_x dx$$

Non-parametric Method: Introduction

- Notation: $x_i; y_i$ denotes **random variable**; $X_i; Y_i$ denotes **realizations**; $x; y$ denotes **random variables or some value of the random variables**
- Realizations are given (sample), they are NOT random in our context

$$\int_{\mathcal{D}_x} x \int_{\mathcal{D}_y} \sum_{i=1}^n X_i dx < \int_{\mathcal{D}_x} \sum_{i=1}^n X_i \int_{\mathcal{D}_y} dx$$

Non-parametric Method: Introduction

- Notation: $x_i; y_i$ denotes **random variable**; $X_i; Y_i$ denotes **realizations**; $x; y$ denotes **random variables or some value of the random variables**
- Realizations are given (sample), they are NOT random in our context
 $\int x < \int_i^n X_i dx \quad < \int_i^n X_i \int dx$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E(y_i|x_i) = x_i^2$, we consider this CEF **point by point**
- That is, estimate $E(y_i|x_i)$ for each possible point of $x_i = x$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E[y_i | x_i] = \alpha x_i^2$, we consider this CEF **point by point**
- That is, estimate $E[y_i | x_i]$ for each possible point of $x_i = x$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E[y_i | x_i] = x_i^2$, we consider this CEF **point by point**
- That is, estimate $E[y_i | x_i]$ for each possible point of $x_i = x$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E[y_i | x_i] = \beta_0 + \beta_1 x_i$, we consider this CEF **point by point**
- That is, estimate $E[y_i | x_i]$ for each possible point of $x_i = x$

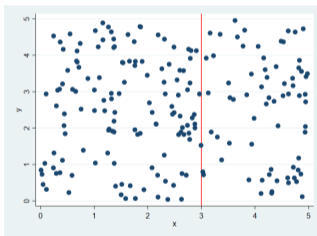
Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E(y_i | x_i) = \beta_0 + \beta_1 x_i$, we consider this CEF **point by point**
- That is, estimate $E(y_i | x_i)$ for each possible point of $x_i = x$

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

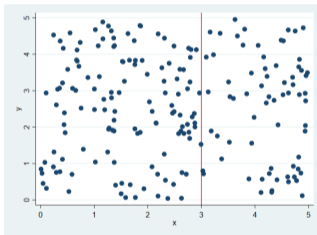


- What is the CDF at $x = 3$? $\hat{F}(x = 3) = ?$
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

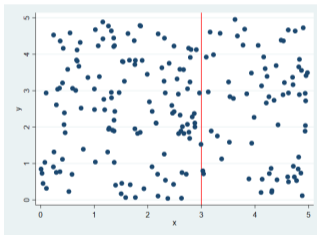


- What is the CDF at $x = 3$? $\hat{F}(x = 3) = ?$
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

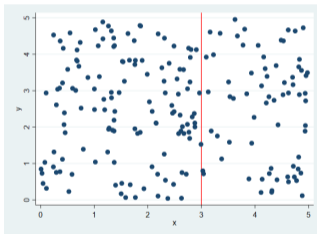


- What is the CDF at $x = 3$? $\hat{F}(x = 3) = ?$
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

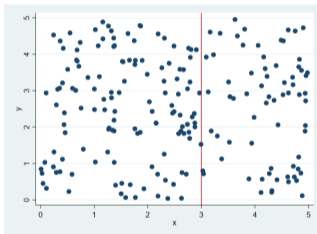


- What is the CDF at $x = 3$? $\hat{F}(x = 3) = ?$
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)



- What is the CDF at $x = 3$? $\hat{F}(x = 3) = ?$
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F}(x) = 3 \quad \frac{1}{n} = \frac{1}{10} \times 3$$

- In general, we have an estimation of $F(x)$ as:

$$F(x) = P(X \leq x) \quad \hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$$

- In general, we have an estimation of $F(x)$ as:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x)$$

- In general, we have an estimation of $F(x)$ as:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x)$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$$

- In general, we have an estimation of $F(x)$ as:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$
$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$
$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$
$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$
$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$
$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$
$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$
$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f(x)$ at some value x as:

$$\hat{f}(x) = \frac{1}{2h} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in [x-h, x+h]} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in [x-h, x+h]}$$
$$\frac{1}{n} = \frac{1}{2h} \mathbb{1}_{x-h \leq X_i \leq x+h}$$

- How to interpret this?
- We count the number of obs within a small interval around x , dividing by the length and the total number of obs
- $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x-h \leq X_i \leq x+h}$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f(x)$ at some value x as:

$$\hat{f}(x) = \frac{1}{2h} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$$

$$\frac{1}{n} = \frac{1}{2h} \mathbf{1}_{x-h \leq X_i \leq x+h}$$

- How to interpret this?
- We count the number of obs within a small interval around x , dividing by the length and the total number of obs
- $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{x-h \leq X_i \leq x+h}$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f(x)$ at some value x as:

$$\hat{f}(x) = \frac{1}{2h} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$$
$$\frac{1}{n} = \frac{1}{2h} \mathbf{1}_{x-h \leq X_i \leq x+h}$$

- How to interpret this?
 - We count the number of obs within a small interval around x , dividing by the length and the total number of obs
 - $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{x-h \leq X_i \leq x+h}$ is the number of obs per unit length
 - When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f(x)$ at some value x as:

$$\hat{f}(x) = \frac{1}{2h} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$$
$$\frac{1}{n} = \frac{1}{2h} \mathbf{1}_{x-h \leq X_i \leq x+h}$$

- How to interpret this?
- We count the number of obs within a small interval around x , dividing by the length and the total number of obs

- $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{x-h \leq X_i \leq x+h}$ is the number of obs per unit length

- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f(x)$ at some value x as:

$$\hat{f}(x) = \frac{1}{2h} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} \frac{1}{2h}$$
$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} = \frac{1}{2h} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$$

- How to interpret this?
- We count the number of obs within a small interval around x , dividing by the length and the total number of obs
- $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

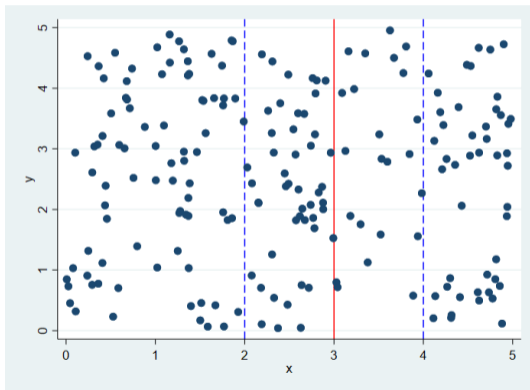
- Then we can write the probability density $f(x)$ at some value x as:

$$\hat{f}(x) = \frac{1}{2h} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$$

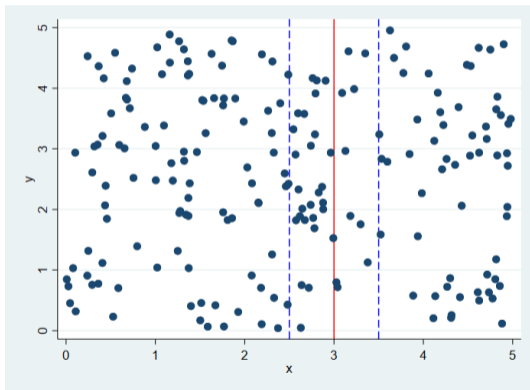
$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]} = \frac{1}{2h} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$$

- How to interpret this?
- We count the number of obs within a small interval around x , dividing by the length and the total number of obs
- $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \in [x-h, x+h]}$ is the number of obs per unit length
- When n is large, we can choose very small h

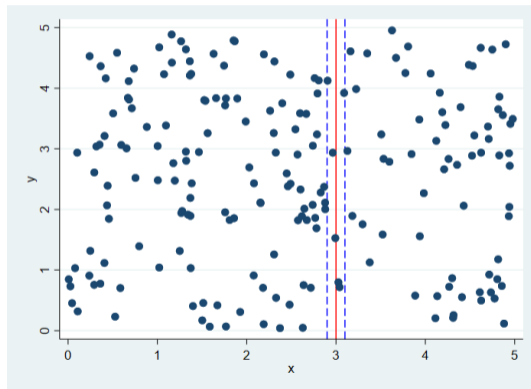
Non-parametric Method: Kernel Regression



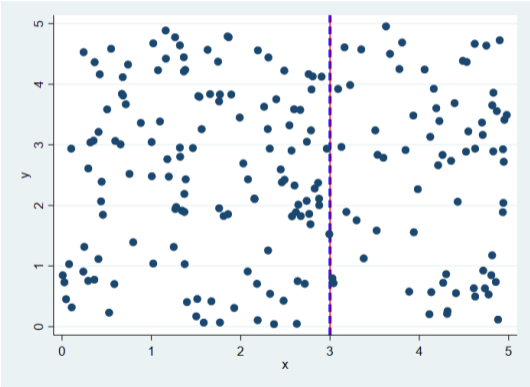
Non-parametric Method: Kernel Regression



Non-parametric Method: Kernel Regression



Non-parametric Method: Kernel Regression



Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2} \mathbb{1}(|v| \leq 1)$. Then we have:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2} \mathbf{1}_{\|v\| \leq 1}$. Then we have:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2} \mathbf{1}_{\|v\| \leq 1}$. Then we have:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2} \mathbf{1}_{\|v\| \leq 1}$. Then we have:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2} \mathbf{1}_{\|v\| \leq 1}$. Then we have:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2} \mathbf{1}_{\|v\| \leq 1}$. Then we have:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(x, x)$ is integrated to 1
 - $k(x, y)$ is symmetric with $k(y, x)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|) \mathbb{1}_{|v| \leq 1}$
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2) \mathbb{1}_{|v| \leq 1}$
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$ & 1
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$ & 1
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$ & 1
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$ & 1
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$ & 1
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$ & 1
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$ & 1
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$ & 1
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$ & 1
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$ & 1
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$ & 1
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$ & 1
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
 - $k(v)$ is integrated to 1
 - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric to the left and to the right
- Triangular Kernel: $k(v) = \frac{1}{2} (1 - |v|)$ & 1
- Epanechnikov Kernel: $k(v) = \frac{3}{4} (1 - v^2)$ & 1
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

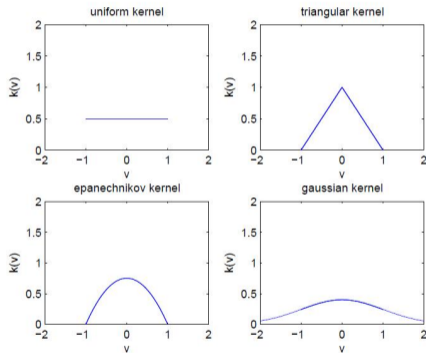


Figure 1: Various Kernels

Non-parametric Method: Kernel Regression

- For multivariate case, let $v = (v_1; v_2; \dots; v_q)$.
- Define product kernel: $K(v) = k(v_1)k(v_2) \dots k(v_q)$.
- The estimator becomes:

$$\hat{f}(x) = \frac{1}{nh_1 h_2 \dots h_q} \sum_i K\left(\frac{X_i - x}{h}\right)$$

- $h = (h_1; h_2; \dots; h_q)$

Non-parametric Method: Kernel Regression

- For multivariate case, let $v = (v_1; v_2; \dots; v_q)$.
- Define product kernel: $K(v) = k(v_1)k(v_2) \dots k(v_q)$.
- The estimator becomes:

$$\hat{f}(x) = \frac{1}{nh_1 h_2 \dots h_q} \sum_i K\left(\frac{X_i - x}{h}\right)$$

- $h = (h_1; h_2; \dots; h_q)$

Non-parametric Method: Kernel Regression

- For multivariate case, let $v = (v_1; v_2; \dots; v_q)$.
- Define product kernel: $K(v) = k(v_1)k(v_2) \dots k(v_q)$.
- The estimator becomes:

$$\hat{f}(x) = \frac{1}{nh_1 h_2 \dots h_q} \sum_i K\left(\frac{X_i - x}{h}\right)$$

- $h = (h_1; h_2; \dots; h_q)$

Non-parametric Method: Kernel Regression

- For multivariate case, let $v = (v_1; v_2; \dots; v_q)$.
- Define product kernel: $K(v) = k(v_1)k(v_2) \dots k(v_q)$.
- The estimator becomes:

$$\hat{f}(x) = \frac{1}{nh_1 h_2 \dots h_q} \sum_i K\left(\frac{X_i - x}{h}\right)$$

- $h = (h_1; h_2; \dots; h_q)$

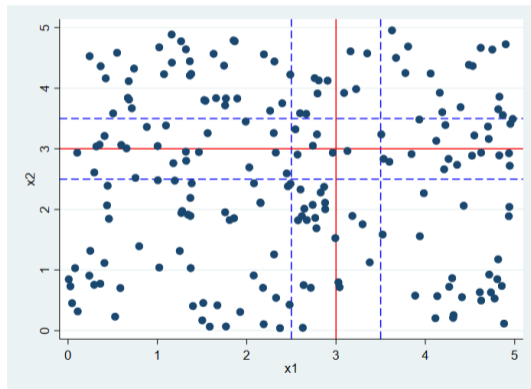
Non-parametric Method: Kernel Regression

- For multivariate case, let $v = (v_1; v_2; \dots; v_q)$.
- Define product kernel: $K(v) = k(v_1)k(v_2) \dots k(v_q)$.
- The estimator becomes:

$$\hat{f}(x) = \frac{1}{nh_1 h_2 \dots h_q} \sum_i K\left(\frac{X_i - x}{h}\right)$$

- $h = (h_1; h_2; \dots; h_q)$

Non-parametric Method: Kernel Regression



Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF **point by point**
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF **point by point**
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF **point by point**
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF **point by point**
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF **point by point**
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF **point by point**
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF **point by point**
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad \text{where} \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{h}$$

- Intuition: The conditional Expectation of Y given $X=x$ is estimated as a **weighted average of observed Y_i closely around x** (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad \text{where} \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{h}$$

- Intuition: The conditional Expectation of Y given $X=x$ is estimated as a **weighted average of observed Y_i closely around x** (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad \text{where} \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{h}$$

- Intuition: The conditional Expectation of Y given $X=x$ is estimated as a **weighted average of observed Y_i closely around x** (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad \text{where} \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{h}$$

- Intuition: The conditional Expectation of Y given $X=x$ is estimated as a **weighted average of observed Y_i closely around x** (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad \text{where} \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{h}$$

- Intuition: The conditional Expectation of Y given $X=x$ is estimated as a **weighted average of observed Y_i closely around x** (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad \text{where} \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{h}$$

- Intuition: The conditional Expectation of Y given $X=x$ is estimated as a **weighted average of observed Y_i closely around x** (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

- We have $g(x) = E(Y|X=x)$ as CEF and $f(x)$ as density for x

Theorem (Asymptotics for N-W Estimator)

Under some regularity conditions, as $n \rightarrow \infty, h_s \rightarrow 0, s=1, \dots, q, nh_1 \dots h_q \rightarrow \infty$ and $nh_1 \dots h_q < \frac{1}{s-1} h_s^6 \rightarrow 0$, we have:

$$\frac{\bar{O}}{nh_1 \dots h_q} \hat{g}(x) - g(x) = \frac{1}{h_s^2} B_s(x) + o_p\left(\frac{1}{n} \int \frac{k(v)^2}{f(x)} dv\right)$$

$$\text{where } B_s(x) = \frac{1}{2f(x)} \frac{d^2 k(v)}{dv^2} \Big|_{v=0} + 2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2}$$

Non-parametric Method: Kernel Regression

- We have $g(x) = E(Y|X)$ as CEF and $f(x)$ as density for x

Theorem (Asymptotics for N-W Estimator)

Under some regularity conditions, as $n \rightarrow \infty, h_s \rightarrow 0, s = 1, \dots, q, nh_1 \dots h_q \rightarrow \infty$ and $nh_1 \dots h_q < \frac{q}{s-1} h_s^6 \rightarrow 0$, we have:

$$\tilde{O}_{nh_1 \dots h_q}(\hat{g}(x) - g(x)) = h_s^2 B_s(x) + N(0; \frac{2}{f(x)} E[k v^2] dv^q)$$

$$\text{where } B_s(x) = \frac{D v^2 k v dv}{2f(x)} - 2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2}$$

Non-parametric Method: Kernel Regression

$$\text{Asymptotic Bias} \sim h_s^q \frac{D \int v^2 k(v) dv}{2f(x)} \left[2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2} \right]$$

$$\text{Asymptotic Variance} \sim \frac{1}{nh_1 \dots h_q} \frac{1}{f(x)} \int D k(v)^2 dv \sim h^{-q}$$

- (1) h_s Bias ; Variance
we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $\sim \int v^2 k(v) dv$; Variance $\sim \int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} ; \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

$$\text{Asymptotic Bias} \sim h_s^q \frac{D \int v^2 k(v) dv}{2f(x)} \left[2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2} \right]$$

$$\text{Asymptotic Variance} \sim \frac{1}{nh_1 \dots h_q} \frac{D \int v^2 k(v)^2 dv}{f(x)^2} h_s^{-q}$$

- (1) h_s Bias ; Variance
we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $\sim D \int v^2 k(v) dv$; Variance $\sim D \int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} ; \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

$$\text{Asymptotic Bias} \sim h_s^q \frac{D^2 v^2 k v dv}{2 f x} \left(2 \frac{\partial f x}{\partial x_s} \frac{\partial g x}{\partial x_s} - f x \frac{\partial^2 g x}{\partial x_s^2} \right)$$

$$\text{Asymptotic Variance} \sim \frac{1}{n h_1 \dots h_q} \frac{x}{f x} D^2 k v^2 dv^q$$

- (1) h_s Bias ; Variance
we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $D^2 v^2 k v dv$; Variance $D^2 k v^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f x}{\partial x_s} \frac{\partial g x}{\partial x_s} ; \frac{\partial^2 g x}{\partial x_s^2}$
- (5) $f x$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

$$\text{Asymptotic Bias} \sim h_s^q \frac{\int v^2 k(v) dv}{2f(x)} \left(2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2} \right)$$

$$\text{Asymptotic Variance} \sim \frac{1}{nh_1 \dots h_q} \frac{\int v^2 k(v)^2 dv}{f(x)^2}$$

- (1) h_s Bias ; Variance
we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $\int v^2 k(v) dv$; Variance $\int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} ; \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

$$\text{Asymptotic Bias} \sim h_s^q \frac{\int v^2 k(v) dv}{2f(x)} \left[2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2} \right]$$

$$\text{Asymptotic Variance} \sim \frac{1}{nh_1 \dots h_q} \frac{\int v^2 k(v)^2 dv}{f(x)^2} \sim h_s^{-q}$$

- (1) h_s Bias ; Variance
we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $\int v^2 k(v) dv$; Variance $\int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} ; \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

$$\text{Asymptotic Bias} \sim h_s^q \frac{D \int v^2 k(v) dv}{2f(x)} \left[2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2} \right]$$

$$\text{Asymptotic Variance} \sim \frac{1}{nh_1 \dots h_q} \frac{D \int v^2 k(v)^2 dv}{f(x)^2} h_s^q$$

- (1) h_s Bias ; Variance
we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $D \int v^2 k(v) dv$; Variance $D \int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} ; \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

$$\text{Asymptotic Bias} \sim h_s^q \frac{D \int v^2 k(v) dv}{2f(x)} \left[2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} - f(x) \frac{\partial^2 g(x)}{\partial x_s^2} \right]$$

$$\text{Asymptotic Variance} \sim \frac{1}{nh_1 \dots h_q} \frac{D \int k(v)^2 dv}{f(x)^2}$$

- (1) h_s Bias ; Variance
we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $D \int v^2 k(v) dv$; Variance $D \int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} ; \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use **piece-wise polynomial** (linear) function to fit data interval by interval

Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use **piece-wise** polynomial (linear) function to fit data interval by interval

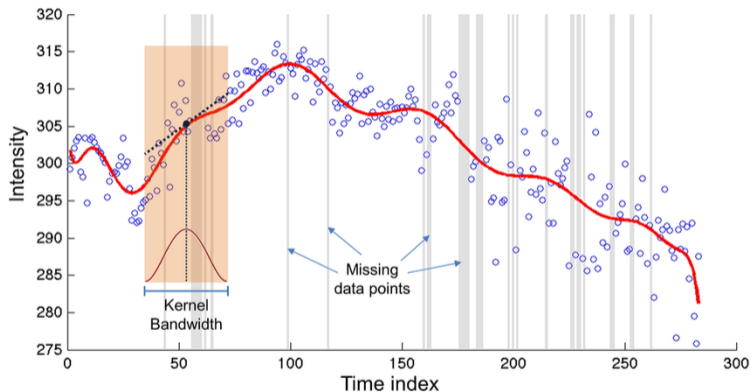
Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use **piece-wise** polynomial (linear) function to fit data interval by interval

Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use **piece-wise** polynomial (linear) function to fit data interval by interval

Non-parametric Method: Local Polynomial



For some $X = x$, we fit $g(x)$ by choosing samples very close to x . Then we fit a polynomial for these observations. (Here, linear)

Non-parametric Method: Local Polynomial

- For $g(x)$, we solve the following optimization problem at each point x :

$$\min_{b_0, b_1, \dots, b_p} \sum_{j=1}^n K \frac{X_j - x}{h} Y_j - b_0 - b_1 (X_j - x) - b_2 (X_j - x)^2 - \dots - b_p (X_j - x)^{p-2}$$

- When $p = 1$, we call it local linear regression
- When $p = 2$, we call it local quadratic regression

Non-parametric Method: Local Polynomial

- For $g(x)$, we solve the following optimization problem at each point x :

$$\min_{b_0, b_1, \dots, b_p} \sum_{i=1}^n \left(\frac{X_i - x}{h} \right)^p \left(Y_i - b_0 - b_1 \frac{X_i - x}{h} - b_2 \left(\frac{X_i - x}{h} \right)^2 - \dots - b_p \left(\frac{X_i - x}{h} \right)^{p-1} \right)^2$$

- When $p = 1$, we call it local linear regression
- When $p = 2$, we call it local quadratic regression

Non-parametric Method: Local Polynomial

- For $g(x)$, we solve the following optimization problem at each point x :

$$\min_{b_0, b_1, \dots, b_p} \sum_{i=1}^n \left(\frac{X_i - x}{h} \right)^{p+1} \left(Y_i - b_0 - b_1 \frac{X_i - x}{h} - b_2 \left(\frac{X_i - x}{h} \right)^2 - \dots - b_p \left(\frac{X_i - x}{h} \right)^p \right)^2$$

- When $p = 1$, we call it local linear regression
- When $p = 2$, we call it local quadratic regression

Non-parametric Method: Local Polynomial

- For $g(x)$, we solve the following optimization problem at each point x :

$$\min_{b_0, b_1, \dots, b_p} \sum_{i=1}^n \left(\frac{X_i - x}{h} \right)^p \left(Y_i - b_0 - b_1 \frac{X_i - x}{h} - b_2 \left(\frac{X_i - x}{h} \right)^2 - \dots - b_p \left(\frac{X_i - x}{h} \right)^p \right)^2$$

- When $p = 1$, we call it local linear regression
- When $p = 2$, we call it local quadratic regression

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
 - Computational burden is large
 - Hard to include information or restrictions over functional form
 - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
 - Computational burden is large
 - Hard to include information or restriction over functional form
 - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
 - Computational burden is large
 - Hard to include information or restriction over functional form
 - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
 - Computational burden is large
 - Hard to include information or restriction over functional form
 - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
 - Computational burden is large
 - Hard to include information or restriction over functional form
 - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
 - Computational burden is large
 - Hard to include information or restriction over functional form
 - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
 - Computational burden is large
 - Hard to include information or restriction over functional form
 - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- As usual, we have a CEF model:

$$Y = g(X) + u$$
$$g(X) = E(Y|X)$$

- We expand the CEF by Taylor Series at zero:

$$g(X) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} X^k$$

Non-parametric Method: Series-based Methods

- As usual, we have a CEF model:

$$Y = g(X) + u$$
$$g(X) = E(Y|X)$$

- We expand the CEF by Taylor Series at zero:

$$g(X) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} X^k$$

Non-parametric Method: Series-based Methods

- As usual, we have a CEF model:

$$Y = g(X) + u$$
$$g(X) = E(Y|X)$$

- We expand the CEF by Taylor Series at zero:

$$g(X) \approx \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} X^k$$

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order global polynomial:

$$g(X) = \sum_{k=0}^K p_k X^k$$

$p_0 X^0 + p_1 X^1 + p_2 X^2 + \dots + p_K X^K$

- We can use OLS to estimate this polynomial
- The vector of $\{p_0; p_1; p_2; \dots; p_K\}$ is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order **global polynomial**:

$$g(X) = \sum_{k=0}^K p_k X^k$$

p_0 x^0 1 ; p_1 x^1 x ; p_2 x^2 x^2 ; \dots ; p_K x^K x^K

- We can use OLS to estimate this polynomial
- The vector of $\{p_0; p_1; p_2; \dots; p_K\}$ is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order **global polynomial**:

$$g(X) = \sum_{k=0}^K p_k X^k$$

$p_0 X^0 = 1; p_1 X^1 = X; p_2 X^2 = X^2; \dots; p_K X^K = X^K$

- We can use OLS to estimate this polynomial
- The vector of $\{p_0; p_1; p_2; \dots; p_K\}$ is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order **global polynomial**:

$$g(X) = \sum_{k=0}^K p_k X^k$$

$p_0 X^0 = 1; p_1 X^1 = X; p_2 X^2 = X^2; \dots; p_K X^K = X^K$

- We can use OLS to estimate this polynomial
- The vector of $\{p_0; p_1; p_2; \dots; p_K\}$ is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order **global polynomial**:

$$g(X) = \sum_{k=0}^K p_k X^k$$

$p_0 X^0 = 1; p_1 X^1 = X; p_2 X^2 = X^2; \dots; p_K X^K = X^K$

- We can use OLS to estimate this polynomial
- The vector of $\{p_0; p_1; p_2; \dots; p_K\}$ is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
$$\{1; x_1; x_2; x_1 x_2; x_1^2; x_2^2; x_1 x_2^2; x_1^2 x_2; x_1^2 x_2^2; \dots\}$$
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
 $1; x_1; x_2; x_1 x_2; x_1^2; x_2^2; x_1 x_2^2; x_1^2 x_2; x_1^2 x_2^2 \dots$
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
 $1; x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 \dots$
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
 $1; x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 \dots$
- Polynomial series has several problems
 - It is very sensitive to outliers
 - The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

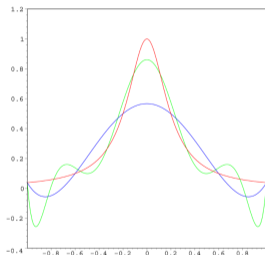
- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
 $1; x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 \dots$
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
 $1; x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 \dots$
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

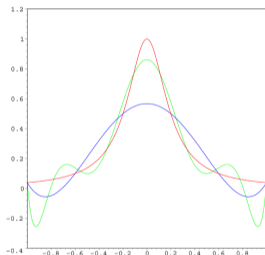
- Runge's phenomenon
- Red: original function; Blue: fifth-order poly; Green: ninth-order poly



- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

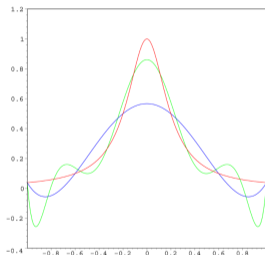
- Runge's phenomenon
- Red: original function; Blue: fifth-order poly; Green: ninth-order poly



- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

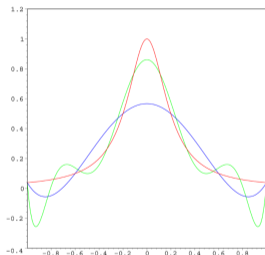
- Runge's phenomenon
- Red: original function; Blue: fifth-order poly; Green: ninth-order poly



- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

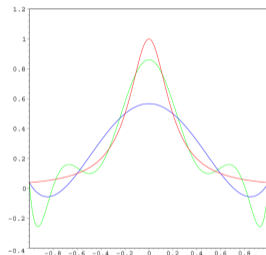
- Runge's phenomenon
- Red: original function; Blue: fifth-order poly; Green: ninth-order poly



- Since the power polynomials are forced to vary somewhere
 - It may be pushed to the boundary
 - The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

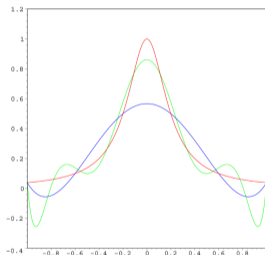
- Runge's phenomenon
- Red: original function; Blue: fifth-order poly; Green: ninth-order poly



- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- Runge's phenomenon
- Red: original function; Blue: fifth-order poly; Green: ninth-order poly



- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion
- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion
- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion

- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion

- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
 - Such as Spline basis and Wavelet basis
 - They are complicated, rarely seen in Applied works
 - But Carol claims that Spline basis is in general a better choice
 - If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y = X\beta + g(Z) + u; \quad E(u|X;Z) = 0; \quad \text{Var}(u|X;Z) = \sigma^2$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y = X^\top \beta + g(Z) + u; \quad E(u|X;Z) = 0; \quad \text{Var}(u|X;Z) = \sigma^2$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y = X^\top \beta + g(Z) + u; \quad E(u|X;Z) = 0; \quad \text{Var}(u|X;Z) = \sigma^2$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y = X^\top \beta + g(Z) + u; \quad E(u|X;Z) = 0; \quad \text{Var}(u|X;Z) = \sigma^2$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$Y - E(Y|Z) = X\beta + E(X|Z) - g(Z) + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $E(Y|Z)$ and $E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$Y - E(Y|Z) = X\beta + g(Z) + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y - E(Y|Z)$ and $X - E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$Y - E(Y|Z) = X\beta - E(X|Z) + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y - E(Y|Z)$ and $X - E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$Y = E(Y|Z) + X \beta + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $E(Y|Z)$ and $E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$Y - E(Y|Z) = X\beta + g(Z) + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y - E(Y|Z)$ and $X - E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$Y - E(Y|Z) = X\beta + g(Z) + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y - E(Y|Z)$ and $X - E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$Y - E(Y|Z) = X\beta + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y - E(Y|Z)$ and $X - E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- In the second step, we subtract $\hat{\alpha}^T$ from Y :

$$Y - X^T \hat{\alpha} = g(Z) + u$$

- $g(Z)$ can be estimated using methods introduced previously

Non-parametric Method: Semi-parametric Model

- In the second step, we subtract $X^\top \beta$ from Y :

$$Y - X^\top \beta = g(Z) + u$$

- $g(Z)$ can be estimated using methods introduced previously

Non-parametric Method: Semi-parametric Model

- In the second step, we subtract X^\top from Y :

$$Y - X^\top = g(Z) + u$$

- $g(Z)$ can be estimated using methods introduced previously

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of \hat{Z} ?
- Can we use the variance from the non-parametric regression directly?
- No! Because $\hat{Y} = \hat{X}^T$ is also estimated
- It contains more uncertainty from the first step
- We need bootstrap

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of \hat{Z} ?
- Can we use the variance from the non-parametric regression directly?
- No! Because $\hat{Y} = \hat{X}^\top$ is also estimated
- It contains more uncertainty from the first step
- We need bootstrap

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of \hat{Z} ?
- Can we use the variance from the non-parametric regression directly?
 - No! Because $\hat{Y} = \hat{X}^\top$ is also estimated
 - It contains more uncertainty from the first step
 - We need bootstrap

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of \hat{Z} ?
- Can we use the variance from the non-parametric regression directly?
- No! Because $\hat{Y} \mid X^{-1}$ is also estimated
- It contains more uncertainty from the first step
- We need bootstrap

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of \hat{Z} ?
- Can we use the variance from the non-parametric regression directly?
- No! Because $\hat{Y} = \hat{X}^\top$ is also estimated
- It contains more uncertainty from the first step
- We need bootstrap

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of $\hat{\beta}$?
- Can we use the variance from the non-parametric regression directly?
- No! Because $\hat{Y} = \hat{X}^T$ is also estimated
- It contains more uncertainty from the first step
- We need bootstrap

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n , draw R new samples of size n , with replacement. Index each new sampler by r
- Step 2: Calculate the simulated variance $\hat{\sigma}_{fx}^2$ by:
$$\hat{\sigma}_{fx}^2 = \frac{1}{R-1} \sum_{r=1}^R (\hat{g}_r(x) - \hat{g}(x))^2$$
- Step 3: Use $\hat{\sigma}_{fx}^2$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n , draw R new samples of size n , with replacement. Index each new sampler by r
- Step 2: Calculate the simulated variance $\hat{\sigma}_x^2$ by:
$$\hat{\sigma}_x^2 = \frac{1}{R-1} \sum_{r=1}^R (\hat{g}_r(x) - \hat{g}(x))^2$$
- Step 3: Use $\hat{\sigma}_x^2$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n , draw R new samples of size n , with replacement. Index each new sampler by

- Step 2: Calculate the simulated variance $\hat{\sigma}_b^2$ by:

$$\hat{\sigma}_b^2 = \frac{1}{R-1} \sum_{r=1}^R (\hat{g}_r - \bar{g})^2$$

- Step 3: Use $\hat{\sigma}_b^2$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n , draw R new samples of size n , with replacement. Index each new sampler by r
- Step 2: Calculate the simulated variance $\hat{\sigma}_x^2$ by:
$$\hat{\sigma}_x^2 = \frac{1}{R-1} \sum_{r=1}^R (\hat{g}_r(x) - \hat{g}(x))^2$$
- Step 3: Use $\hat{\sigma}_x^2$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n , draw R new samples of size n , with replacement. Index each new sampler by r
- Step 2: Calculate the simulated variance $\hat{\sigma}_b^2$ by:
$$\hat{\sigma}_b^2 = \frac{1}{R-1} \sum_{r=1}^R (\hat{g}_r - \bar{g})^2$$
- Step 3: Use $\hat{\sigma}_b^2$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{\Lambda}^1, \hat{\Lambda}^2, \dots, \hat{\Lambda}^R$
- We have an empirical distribution of $\hat{\Lambda}$
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^-, q_{1-\alpha}^-$
- q_{α}^- is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{x}^1; \hat{x}^2; \dots; \hat{x}^R$
- We have an empirical distribution of \hat{x}
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^{\sim}; q_{1-\alpha}^{\sim}$
- q_{α}^{\sim} is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{\theta}^1, \hat{\theta}^2, \dots, \hat{\theta}^R$
- We have an empirical distribution of $\hat{\theta}^r$
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^{\sim}; q_{1-\alpha}^{\sim}$
- q_{α}^{\sim} is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{x}^1; \hat{x}^2; \dots; \hat{x}^R$
- We have an empirical distribution of \hat{x}^r
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^{\sim}; q_{1-\alpha}^{\sim}$
- q_{α}^{\sim} is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{x}^1; \hat{x}^2; \dots; \hat{x}^R$
- We have an empirical distribution of \hat{x}^r
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^{\sim}; q_{1-\alpha}^{\sim}$
- q_{α}^{\sim} is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{x}^1; \hat{x}^2; \dots; \hat{x}^R$
- We have an empirical distribution of \hat{x}^r
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^{\sim}; q_{1-\alpha}^{\sim}$
- q_{α}^{\sim} is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{\theta}^1; \hat{\theta}^2; \dots; \hat{\theta}^R$
- We have an empirical distribution of $\hat{\theta}^r$
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^{\sim}; q_{1-\alpha}^{\sim}$
- q_{α}^{\sim} is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better choice is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{\theta}^1; \hat{\theta}^2; \dots; \hat{\theta}^R$
- We have an empirical distribution of $\hat{\theta}^r$
- The bootstrap 100(1 - α)% confidence interval is then: $q_{\alpha}^{\sim}; q_{1-\alpha}^{\sim}$
- q_{α}^{\sim} is the quantile of this empirical distribution

Non-parametric Method: Application

- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
 - Causal inference highly depends on non-parametric techniques
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t , but not the structure behind it
 - Predict stock price, machine learning, RDD fitting
- We will show these in the following lectures

Non-parametric Method: Application

- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
 - Causal inference highly depends on non-parametric techniques
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t , but not the structure behind it
 - Predict stock price, machine learning, RDD fitting
- We will show these in the following lectures

Non-parametric Method: Application

- Anything related to estimation of CEF
- **Potential outcome framework is non-parametric**
Causal inference highly depends on non-parametric techniques
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t , but not the structure behind it
Predict stock price, machine learning, RDD fitting
- We will show these in the following lectures

Non-parametric Method: Application

- Anything related to estimation of CEF
- **Potential outcome framework is non-parametric**
Causal inference highly depends on non-parametric techniques
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t , but not the structure behind it
Predict stock price, machine learning, RDD fitting
- We will show these in the following lectures

Non-parametric Method: Application

- Anything related to estimation of CEF
- **Potential outcome framework is non-parametric**
Causal inference highly depends on non-parametric techniques
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t , but not the structure behind it
Predict stock price, machine learning, RDD fitting
- We will show these in the following lectures

Non-parametric Method: Application

- Anything related to estimation of CEF
- **Potential outcome framework is non-parametric**
Causal inference highly depends on non-parametric techniques
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t , but not the structure behind it
Predict stock price, machine learning, RDD fitting
- We will show these in the following lectures

Non-parametric Method: Application

- Paper report
Dube et al. (2020) Monopsony in Online Labor Markets

Non-parametric Method: Application

- Paper report
Dube et al. (2020) Monopsony in Online Labor Markets

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: K-NN estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - May need to incorporate restrictions
 - Requires large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
 - Kernel-based methods: N-W estimator, Local polynomial
 - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful when you want to do prediction, or when you want to implement causal inference in a complicated context
- However, they have weaknesses: **Not always better to make model more flexible**
 - Hard to incorporate restrictions
 - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

References

- Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri. 2020. "Monopsony in Online Labor Markets." *American Economic Review: Insights* 2 (1):33–46.
- Robinson, Peter M. 1988. "Root-N-consistent Semiparametric Regression." *Econometrica* :931–954.