Econ 675 Assignment 1

Nathan Mather

October 12, 2018

Contents

1 Kernal Density Estimation

1.1 Part 1

Start by noting that

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

Now taking the expectation of $\hat{f}^{(s)}(x)$ that we can apply the linearity of expectations to move the expectation inside the sum. Then we can use the i.i.d. assumption to show the sum is just n times the expectation. This leaves us with

$$\mathrm{E}[\hat{f}^{(s)}(x)] = \mathrm{E}\left[\frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{x_i - x}{h}\right)\right] = \int_{-\infty}^{\infty} \frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{z - x}{h}\right) f(z)dz$$

Where the second equality is just by the definition of the expectation. Next we use integration by parts. Note that

$$\int_{-\infty}^{\infty} \frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{z-x}{h}\right) f(z)dz = -\int_{-\infty}^{\infty} \frac{(-1)^s}{h^s} k^{(s-1)}\left(\frac{z-x}{h}\right) f^{(1)}(z)dz$$

Iterating this s times gives us

$$\int_{-\infty}^{\infty} \frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{z-x}{h}\right) f(z)dz = (-1)^s \int_{-\infty}^{\infty} \frac{(-1)^s}{h} k\left(\frac{z-x}{h}\right) f^{(s)}(z)dz = \int_{-\infty}^{\infty} \frac{1}{h} k\left(\frac{z-x}{h}\right) f^{(s)}(z)dz$$

Next we apply change of variables. let $u = \frac{z-x}{h}$ Note that $du = \frac{1}{h}dz$ so we get

$$\int_{-\infty}^{\infty} k(u)f^{(s)}(x+hu)du$$

Next we Taylor expand $f^{(s)}(x+hu)$ to the Pth order about x. Recall from properties of the kernal estimator that $\int_{-\infty}^{\infty} k(u)du = 1$ and that $\int_{-\infty}^{\infty} k(u)u^j du = 0$ for all $j \neq p$ This gives us

$$f^{(s)}(x) + \frac{1}{P!}f^{(s+P)}(x)h^P \int_{-\infty}^{\infty} k(u)u^P du + o(h^P) = f^{(s)}(x) + \frac{1}{P!}f^{(s+P)}(x)h^p\mu_P(k) + o(h^P)$$

which is the desired result.

Now for the variance recall again that

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

So by the i.i.d. assumption we can get that

$$V\left(\hat{f}^{(s)}(x)\right) = \frac{1}{nh^{2+2s}} V\left(k^{(s)}\left(\frac{x_i - x}{h}\right)\right)$$

$$V\left(\hat{f}^{(s)}(x)\right) = \frac{1}{nh^{2+2s}} V\left(k^{(s)}\left(\frac{x_i - x}{h}\right)\right) \tag{1}$$

$$= \frac{1}{n2h^{2+2s}} E\left[\left(k^{(s)}\left(\frac{x_i - x}{h}\right)\right)^2\right] - \frac{1}{nh^{2+2s}} E\left[\left(k^{(s)}\left(\frac{x_i - x}{h}\right)\right)^2\right]^2 \tag{2}$$

$$= \frac{1}{nh^{2+2s}} E\left[\left(k^{(s)}\left(\frac{x_i - x}{h}\right)\right)^2\right] - \frac{1}{n}\left(\frac{1}{h^{1+s}} E\left[\left(k^{(s)}\left(\frac{x_i - x}{h}\right)\right)^2\right]\right)^2 \tag{3}$$

$$= \frac{1}{nh^{2+2s}} \int_{-\infty}^{\infty} k^{(s)}\left(\frac{x_i - x}{h}\right)^2 f(z)dz + \frac{1}{nh^{2+2s}} f^{(n)}(X)^2 \tag{4}$$

$$= \frac{1}{nh^{1+2s}} \int_{-\infty}^{\infty} k^{(s)}(u)^2 f(x+hu)du + o\left(\frac{1}{nh^{2+2s}}\right) \tag{5}$$

$$= \frac{1}{nh^{1+2s}} \cdot \vartheta_s(K) + o\left(\frac{1}{nh^{2+2s}}\right) \tag{6}$$

1.2   part 2

We start with the following AMISE

$$AIMSE[h] = \int \left[\left(h_n^P \cdot \mu_P(K) \cdot \frac{f^{(P+s)}(x)}{P!} \right)^2 + \frac{1}{nh_n^{1+2s}} \cdot \vartheta_s(K) \cdot f(x) \right] dx$$

Using the ϑ notation so $\vartheta_{P+s}(f) = \int (f^{(P+s)}(x))^2$ and recalling that f(x) integrates to 1 we can rewrite this as

$$= h_n^{2P} \left(\frac{\mu_P(K)}{P!} \right)^2 \vartheta_{P+s}(f) + \frac{\vartheta_s(K)}{nh_n^{1+2s}}$$

Now taking first order conditions and solving for h

$$\frac{d}{dh} AIMSE[h] = 2Ph_n^{2p-1} \left(\frac{\mu_P(K)}{P!} \right)^2 \vartheta_{P+s}(f) - (1+2s)\frac{\vartheta_s(K)}{nh_n^{2+2s}} = 0$$

$$\implies 2Ph^{1+2P+2s} \left(\frac{\mu_P(K)}{P!} \right)^2 \vartheta_{P+s}(f) = (1+2s)\frac{\vartheta_s(K)}{n}$$

Thus, we get the AIMSE-optimal bandwidth choice.

$$h_{AIMSE_s} = \left[\frac{(2s+1)(P!)^2}{2P} \frac{\vartheta_s(K)}{\vartheta_{s+P}(f) \cdot \mu_P(K)^2} \frac{1}{n} \right]^{\frac{1}{1+2P+2s}}$$

Least squares cross-validation is a fully automatic data-driven method of selecting the smoothing parameter h. THis is based on the principle of selecting bandwidth that minimizes the integrated squared error of the resulting estimate.The estimate used is

$$\hat{h}_{CV} = \arg\min_h \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \bar{k}\left(\frac{X_i - X_j}{h} \right) - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{j=1,i\neq j}^n k\left(\frac{X_i - X_j}{h} \right)$$

1.3   Monte Carlo experiment

1.3.1   a

First, we want to compute the theoretically optimal bandwidth for $s = 0$, $n = 1000$, using the Epanechnikov kernel $(P = 2)$, with the following Gaussian DGP:

$$x_i \sim 0.5\mathcal{N}(-1.5, -1.5) + 0.5\mathcal{N}(1, 1)$$

Filling in what we know so far we have :

$$h_{AIMSE_s} = \left[\frac{\vartheta_0(K)}{\vartheta_2(f) \cdot \mu_2(K)^2} \frac{1}{1000} \right]^{\frac{1}{5}}$$

So we need the second moment of K and the first moment of the second derivative of k squared. We can get two of these values from the table in Bruce Hanson's nonparametric notes. Giving us.

$$h_{AIMSE_s} = \left[\frac{\frac{3}{5}}{\vartheta_2(f) \cdot \frac{1}{5}^2} \frac{1}{1000} \right]^{\frac{1}{5}}$$

3

The second derivative of the normal density φ with mean μ variance σ^2 is

$$\varphi''_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \left[\left(\frac{(x-\mu)}{\sigma^2} \right)^2 - \frac{1}{\sigma^2} \right]$$

now useing the linearity of integrals we can find $\vartheta_2(f)$

$$\vartheta_2(f) = \int_{-\infty}^{\infty} [0.5\varphi''_{1,1}(x) + 0.5\varphi''_{-1.5,1.5}(x)]^2 dx \approx 0.03883397$$

Where the approximation comes from R

Finally, pluging this in gives the theoretically optimal bandwidth is:

$$h* = 0.8267532$$

1.3.2 b

Below Is the table of $\widehat{IMSE}^{lI}$ results and $\widehat{IMSE}^{LO}$ results by bandwidth h. I only included the table from R. My stata code was significantly slower and so I only ran 10 repetitions. The results were as close as expected.

h	imse_li	imse_lo	d_h_hat
0.5	0.000168	0.000168	1.05
0.6	0.000139	0.00014	1.05
0.7	0.000122	0.000124	1.05
0.8	0.000114	0.000117	1.05
0.9	0.000114	0.000117	1.05
1	0.000121	0.000124	1.05
1.1	0.000135	0.000138	1.05
1.2	0.000156	0.000159	1.05
1.3	0.000184	0.000187	1.05
1.4	0.000218	0.000222	1.05
1.5	0.000259	0.000263	1.05

My graphs from both programs are below
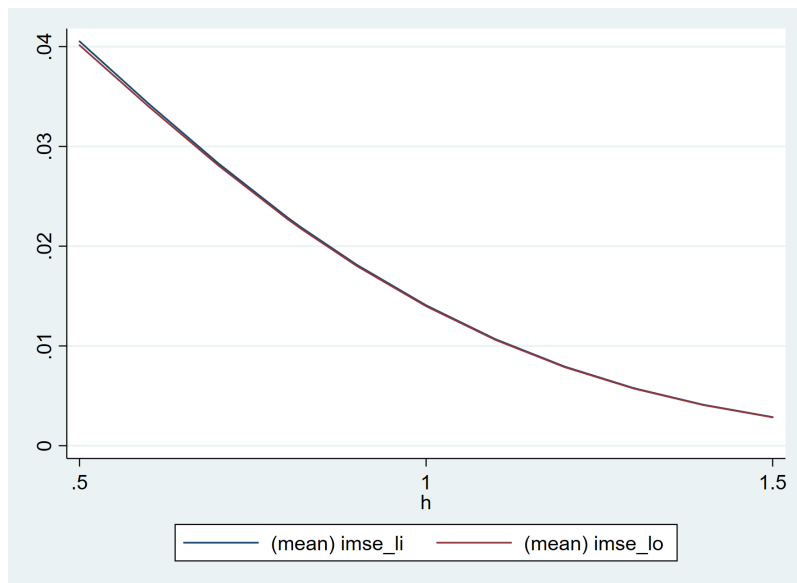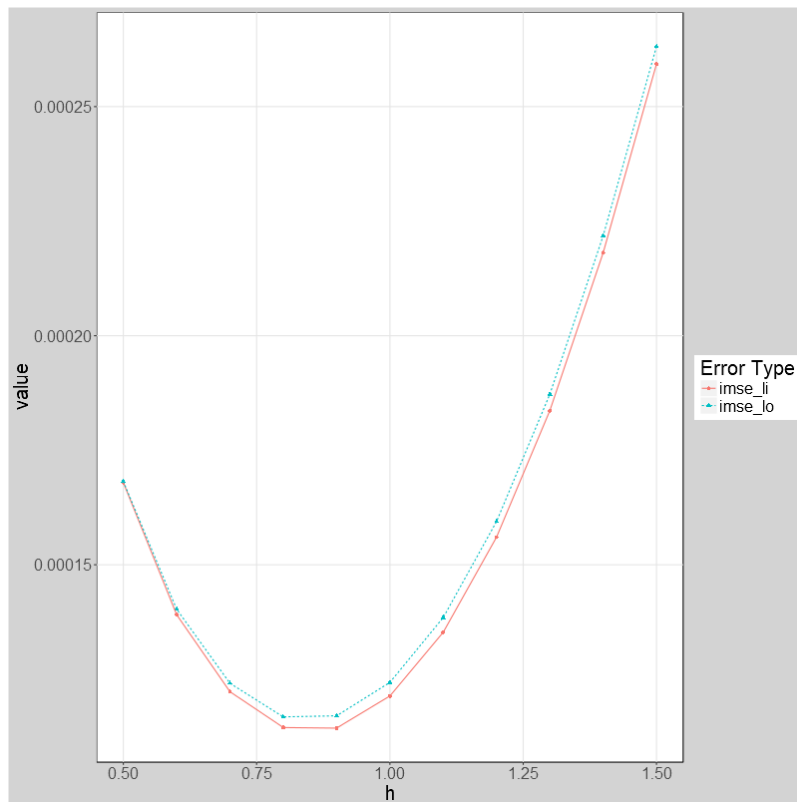
1.3.3   c

Intuitively the difference between the two estimators, LI and LO, is that the LI includes the extra zero term in the sum since we include $x_i - x_i$. As the size of the sample increases this contribution to the overal average will go to zero. Meaning that the LI IMSE will also converge to the correct estimate. s

1.3.4 d

The "d_h_hat" column of the graph in part c is my calculation of this over the 1000 iterations. The value it came up with was 1.04. This is somewhat close but, as expected, not exactly correct.

2 Question 2: Linear Smoothers, Cross-validation and Series

2.1 Part 1

For local polynomial regression we want to estimate $e(x) = \mathrm{E}[y_i|x_i = x]$. The idea of a local polynomial regression is to estimate e(x) around the point x using a polynomial of degree p. We estimate this polynomial with weighted least squares. For a given x, we want to solve.

$$\hat{\boldsymbol{\beta}}_{LP}(x) = \arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} [y_i - \boldsymbol{r}_p(x_i - x)'\boldsymbol{\beta}]^2 K\left(\frac{x_i - x}{h}\right)$$

where $\boldsymbol{r}_p(x) = (1, x, x^2, ..., x^p)'$ and $\hat{e}(x) = \hat{\beta}_0$
from the lecture notes we can get that

$$\hat{\boldsymbol{\beta}}_{LP}(x) = (\boldsymbol{R}_p'\boldsymbol{W}\boldsymbol{R}_p')^{-1}\boldsymbol{R}_p'\boldsymbol{W}\boldsymbol{y}$$

I think This simplifies to the following

$$\hat{e}(x) = e_1'\left(\sum_{i=1}^{n} \boldsymbol{r}_p(x_i - x)\boldsymbol{r}_p(x_i - x)'w_i\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{r}_p(x_i - x)w_i y_i\right)$$

where $wi = K\left(\frac{x_i - x}{h}\right)$
Now for the series estimation.

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{series}} = \arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} (y_i - \boldsymbol{r}_p(x_i)'\boldsymbol{\beta})^2$$

where $\boldsymbol{r}_p(x_i) = (1, x_x, x_i^2, ..., x_i^p)$ and

$$\hat{e}(x) = \boldsymbol{r}_p(x)'\hat{\boldsymbol{B}}_{series}$$

Together we get
$$\hat{\boldsymbol{B}}_{series} = (\boldsymbol{R}_p'\boldsymbol{R}_p)^{-1}\boldsymbol{R}_p\boldsymbol{y}$$

so

$$\hat{e}(x) = \boldsymbol{r}_p(x)'(\boldsymbol{R}_p'\boldsymbol{R}_p)^{-1}\boldsymbol{R}_p\boldsymbol{y}$$

Writing this in linear summation form I believe we get

$$\hat{e}(x) = \boldsymbol{r}_p(x)'\left(\sum_{i=1}^{n} \boldsymbol{r}_p(x_i)\boldsymbol{r}_p(x_i)'\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{r}_p(x_i)y_i\right)$$

2.2 Part 2

We want to choose the tuning parameter to minimize the mean squared leave one out error which is

$$\hat{c} = \arg\min_c \frac{1}{n} \sum_{i=1}^n (y_i - \hat{e}_{(i)}(x_i; c))^2$$

where $\hat{e}_{(i)}(x_i)$ is the estimator of the regression function that leaves out x_i
We can write the local polynomial series estimator as

$$\hat{e}(x) = \boldsymbol{Sy}$$

Where S is the smoothing matrix. Note that the rows of S sum to one so $\boldsymbol{S1} = \boldsymbol{1}$. For the leave one out estimator we want to use $\boldsymbol{S}$ but with the i_{th} row and column removed. If we let the elements of $\boldsymbol{s}$ be denoted by w_{ij} than deleting the i_{th} column means that the i_{th} row will now sum to $1 - w_{ij}$. So, we divide by $1 - w_{ij}$ to renormalize and get the the leave-one-out estimator is

$$\hat{e}_{(i)}(x_i) = \frac{1}{1 - w_{ij}} \sum_{j=1 j \neq i}^n w_{ij} y_i$$

The full sample estimator is

$$\hat{e}(x_i) = \sum_{j=1}^n w_{ij} y_i$$

Together we can get that

$$\hat{e}_{(i)}(x_i)(1 - w_{ij}) = \sum_{j=1 j \neq i}^n w_{ij} y_i$$

$$\implies \hat{e}_{(i)}(x_i) = \sum_{j=1 j \neq i}^n w_{ij} y_i + w_{ij}\hat{e}_{(i)}(x_i) = \sum_{j=1}^n w_{ij} y_i + w_{ij}\hat{e}_{(i)}(x_i) - w_{ij} y_i = \hat{e}_{(i)}(x_i) + w_{ij}\hat{e}_{(i)}(x_i) - w_{ij} y_i$$

$$\implies y - \hat{e}_{(i)}(x_i) = y - \hat{e}_{(i)}(x_i) - w_{ij}\hat{e}_{(i)}(x_i) + w_{ij} y_i$$
$$= y - \hat{e}_{(i)}(x_i) + w_{ij}(y_i - \hat{e}_{(i)}))$$

$$\implies y_i - \hat{e}_{(i)}(x_i) = \frac{1}{1 - w_{ij}}(y_i - \hat{e}(x_i))$$

Which is what we wanted

2.3 part 3

Note that we have iid data and the $\sum_{i=1}^n w_{n,i}(x_i) = 1$ first we want to find

$$\mathrm{E}[\hat{e}(x)|\boldsymbol{x}] = \mathrm{E}\left[\sum_{i=1}^n w_{n,i}(x_i)y_i|\boldsymbol{x}\right] = \sum_{i=1}^n \mathrm{E}\left[w_{n,i}(x_i)y_i|\boldsymbol{x}\right] = \sum_{i=1}^n w_{n,i}(x_i)\mathrm{E}\left[y_i|\boldsymbol{x}\right] = \mathrm{E}[y_i|\boldsymbol{x}]$$

Now as long as we have a bounded second moment we can use CLT to get asymptotic normality. Now to calculate the variance:

$$V[\hat{e}(x)|\boldsymbol{x}] = V\left[\sum_{i=1}^{n} w_{n,i}(x_i)y_i|\boldsymbol{x}\right] = \sum_{i=1}^{n} V\left[w_{n,i}(x_i)y_i|\boldsymbol{x}\right] = \sum_{i=1}^{n} w_{n,i}(x_i)^2 V\left[y_i|\boldsymbol{x}\right]$$

Then if we assume homoroskedasticity we get the estimator

$$\hat{V}(x) = \hat{\sigma}^2 \sum_{i=1}^{n} w_{n,i}(x)^2$$

Where $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{e}(x_i))^2$

2.4 part 4

The pointwise asymptotically valid 95% convidence interval for $e(x)$ is

$$CI_{95}(x) = [\hat{e}(x) - 1.96\sqrt{\hat{V}(x)}, \hat{e}(x) + 1.96\sqrt{\hat{V}(x)}]$$

This is just a confidence interval for a given point. applying this to a grid of points across the line and interpreting that as a band for the function is incorrect. For uniformly valid inference we need that the estimate is less that the cutoff for all values of x, not just one specific x.
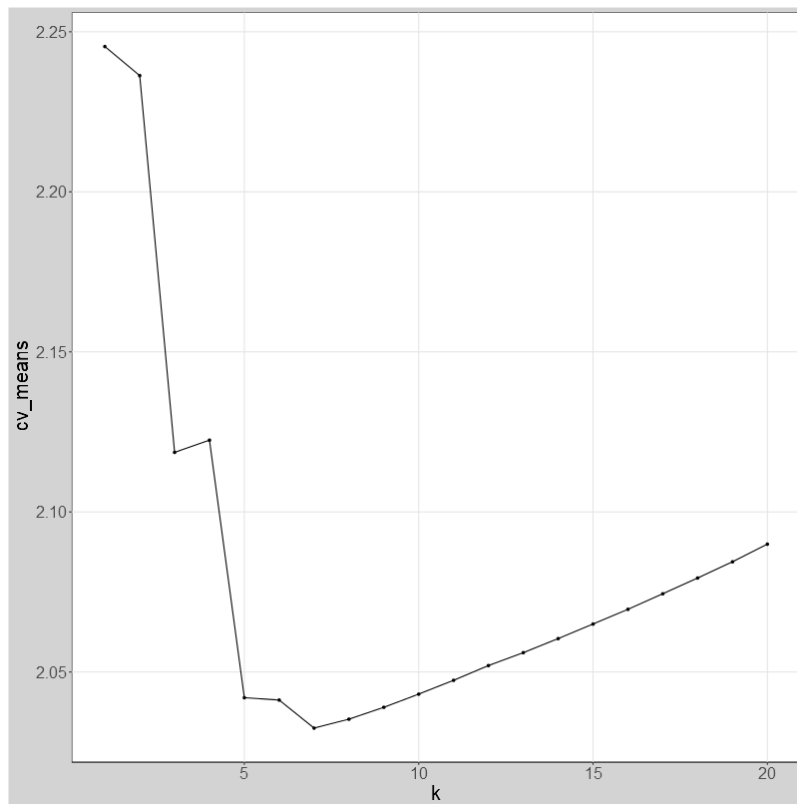
2.5 part 5

2.5.1 a
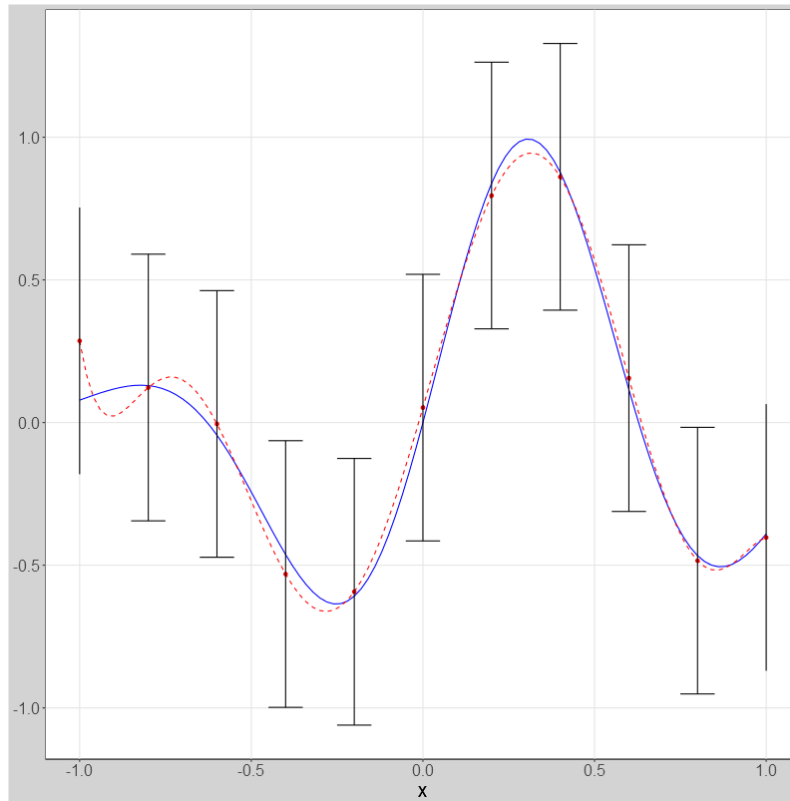
See the code in appendix

2.5.2 B

The plot of the CV(K) simulations is below

2.6 C

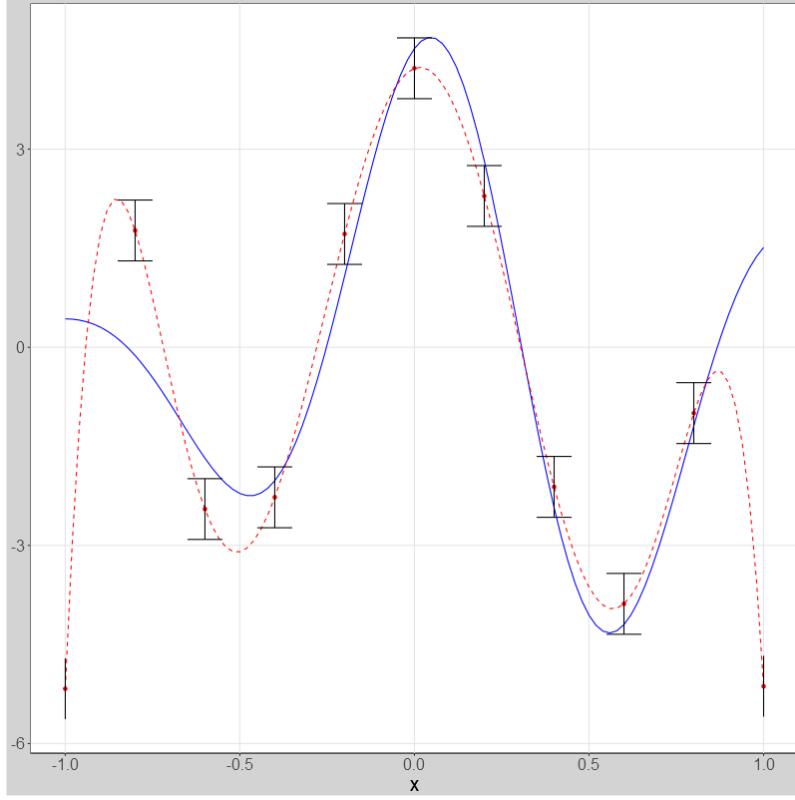My plot is below. I used homoscedastic standard errors

2.7 Part D

Calculating the derivative of $u(x)$ we get

$$e^{(0.1 \cdot (4x-1)^2)}[5 \cdot cos(5x) - 0.8 \cdot (4x-1)sin(5x)]$$

Taking the derivative of the estimated euqation we get

$$\hat{u}'(x) = \beta_1 + 2\beta_2 x + 3\beta_3 x^2 + 4beta_4 x^3 + 5\beta_5 x^4 + 6\beta_6 x^5 + 7\beta_7 x^6$$

I plot the corresponding curves below

3 Question 3: Semiparametric Semi-Linear Model

3.1 part 1

first note that θ_0 cannot contain a constant. since $\alpha + g(z) = [\alpha + c] + [g(z) - c] \equiv \alpha_{new} + g_{new}(z)$ the sum of the new g and intercept are observationally equivalent to the old ones so they cannot be identified. From Li and Racine page 223 we can see that the requirements for identifiability are that $E[(t_i - h_0(x_i))(t_i - h_0(x_i))']$ is positive definite. 3

To prove the moment condition let's start with the expectation of interest and apply the law of iterated expectations.

$$E[(t_i - h_o(x_i))(y_i - t_i\theta_0)] = E[E[(t_i - h_o(x_i))(y_i - t_i\theta_0)|x_i]] = E[E[(t_i - h_o(x_i))(g_0(x_i) + \epsilon_i)|x_i]]$$

$$= E[E[(t_i - h_o(x_i))g_o(x_i)|x_i]] + E[E[(t_i - h_o(x_i))\epsilon_i|x_i]]$$
$$= E[g_o(x_i)E[(t_i - h_o(x_i))|x_i]] + E[(t_i - h_o(x_i))E[\epsilon_i|x_i, i_i]] = 0$$

Now to find a closed form solution for θ_0.

$$E[(t_i - h_o(x_i)y_i)] - E[(t_i - h_o(x_i)t_i)]\theta_0 = 0$$
$$\implies \theta_0 = \frac{E[(t_i - h_o(x_i)y_i)]}{E[(t_i - h_o(x_i)t_i)]}$$

The IV interpretation can be given as follows. Let $yi = t_i\theta_0 + g_0(x_i) + \epsilon_i = t_i\theta_0 + \mu_i$. Now μ_i is uncorrelated with t_i so we can define and instrument $z_i = t_i - h_0(x_i)$ which has the property of $E[z_i\mu_i] = 0$ and $E[t_i z_i] \neq 0$ so it is a valid instrument.

3.2 3.2

3.3 3.3

3.4 3.4

3.4.1 Part a

see code in appendix

3.4.2 part b

)
My results from R are in the table below

K	Theta	Bias	S.D	V_HCO	Rejection rate
6.000	1.841	0.467	3.607	0.462	0.980
11.000	-0.167	0.231	0.082	0.218	0.093
21.000	-0.161	0.232	0.080	0.214	0.093
26.000	-0.162	0.235	0.081	0.214	0.102
56.000	-0.140	0.227	0.071	0.199	0.128
61.000	-0.124	0.221	0.064	0.193	0.122
126.000	0.008	0.113	0.013	0.100	0.094
131.000	0.008	0.114	0.013	0.100	0.093
252.000	0.008	0.128	0.017	0.094	0.149
257.000	0.008	0.130	0.017	0.094	0.153
262.000	0.008	0.132	0.017	0.094	0.154
267.000	0.009	0.132	0.018	0.094	0.162
272.000	0.009	0.133	0.018	0.094	0.162
277.000	0.008	0.135	0.018	0.094	0.170