Solutions Manual to Accompany
Nonparametric Econometrics
(Answers to Odd Numbered Questions)

Theory and Practice

October 10, 2013

Contents

Preface

This monograph contains solutions to the exercises appearing in Li and Racine (2007).

Solutions to the empirical exercises are provided in the `R` environment for statistical computing and graphics (www.r-project.org) and make use of the `np` package (Hayfield and Racine (2008)) which must be loaded prior to running the examples (in `R` type `install.packages("np")` followed by `library("np")`).

Chapter 1

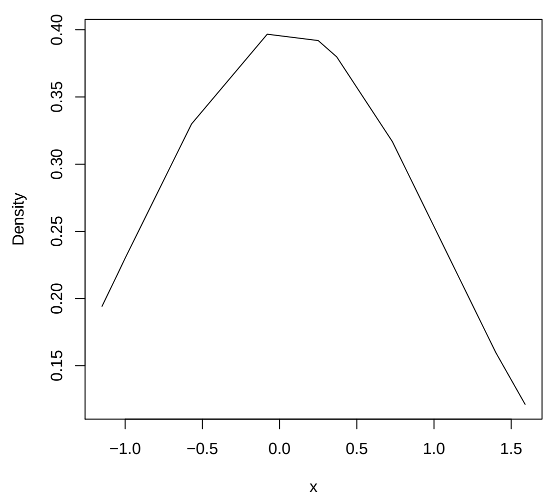
Density Estimation: Solutions

Exercise 1.1. First create (and sort for some of the plots below) the data using R:

```
R> x <- c(-0.57, 0.25, -0.08, 1.40, -1.05, -1.00, 0.37, -1.15, 0.73, 1.59)
R> x <- sort(x)
```

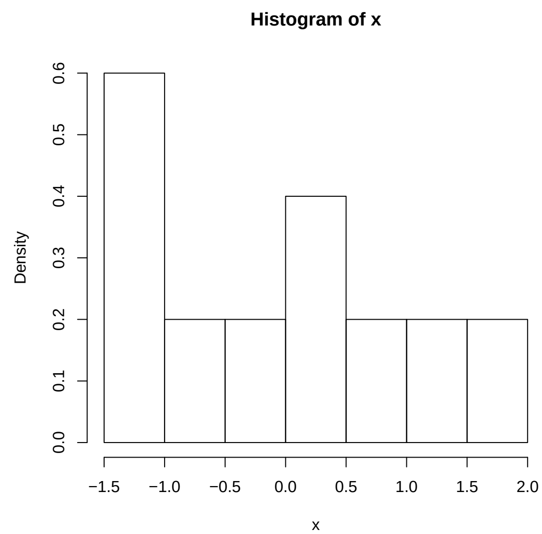
(i) Compute and graph the parametric density function for this data (i.e., compute $\hat{\mu}$ and $\hat{\sigma}^2$) assuming an underlying normal distribution.

```
R> plot(x,dnorm(x,mean=mean(x),sd=sd(x)),ylab="Density",type="l")
```



(ii) Compute and graph a histogram for this data using bin widths of 0.5 ranging from -1.5 through 2.0 (the default values in R).

```
R> hist(x,prob=TRUE)
```



(iii) Using the same tiny sample of data, compute the kernel estimator of the density function for every sample realization using the bandwidth $h = 1.5$ (*we use an Epanechnikov kernel*). Show all steps.

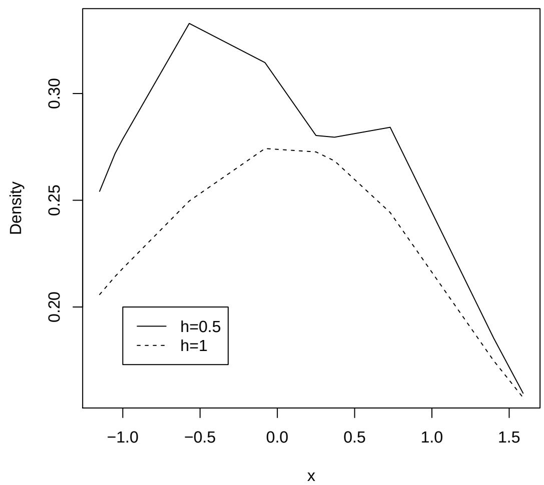
```
R> kernel <- function(x,y,h) {
+   z <- (x-y)/h
+   ifelse(abs(z)<sqrt(5),(1-z^2/5)*(3/(4*sqrt(5))),0)
+ }
R> h <- 1
R> fh1 <- numeric(length(x))
R> for(i in 1:length(x)) {
+   fh1[i] <- sum(kernel(x,x[i],h)/(length(x)*h))
+ }
```

(iv) Using the same data, compute the kernel estimator of the density function for every sample realization using the bandwidth $h = 0.5$. Show all steps.

```
R> kernel <- function(x,y,h) {
+   z <- (x-y)/h
+   ifelse(abs(z)<sqrt(5),(1-z^2/5)*(3/(4*sqrt(5))),0)
+ }
R> h <- 0.5
R> fh05 <- numeric(length(x))
R> for(i in 1:length(x)) {
+   fh05[i] <- sum(kernel(x,x[i],h)/(length(x)*h))
+ }
```

(v) On the same axes, graph your estimates of the density functions using a smooth curve to “connect the dots” for each function.

```
R> plot(x,fh05,type="l",ylab="Density",lty=1)
R> lines(x,fh1,lty=2)
R> legend(-1,0.2,c("h=0.5","h=1"),lty=c(1,2))
```



(vi) Describe the effect of *increasing* the smoothing parameter on the estimated density function.

Increasing the bandwidth results in a 'less rough' density estimate.

Exercise 1.2.

Exercise 1.3.

(i) First we have

$$E(\hat{p}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot nE(X_i) = p,$$

since

$$E(X_i) = \sum_{x_i=1,0} x_i p(x_i) = (1)P(x_i = 1) + (0)P(x_i = 0) = P(x_i = 1) = P(H) = p.$$

Next,

$$Var(\hat{p}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{k=1}^{n} Var(X_i) = \frac{1}{n^2}\cdot nVar(X_i) = \frac{p(1-p)}{n},$$

because

$$Var(X_i) = E\left(X_i^2\right) - \left[E\left(X_i\right)\right]^2 = E\left(X_i\right) - \left[E\left(X_i\right)\right]^2 = p - p^2 = p(1-p),$$

since $X_i^2 \equiv X_i$ and $E(X_i) = p$ as shown above. Therefore,

$$MSE\left(\hat{p}\right) = E\left[\hat{p} - p\right]^2 = Var\left(\hat{p}\right) + \left[bias\left(\hat{p}\right)\right]^2 = \frac{p(1-p)}{n}.$$

(ii) By the Markov inequality (see e.g. Li and Racine (2007, (A.24) on page 690))

$$\mathrm{P}\left\{|\hat{p}-p| \geq \epsilon\right\} \leq \frac{1}{\epsilon^2} E\left[\hat{p}-p\right]^2 = \frac{p(1-p)}{n} \cdot \frac{1}{\epsilon^2} \to 0$$

as $n \to \infty$, since $\epsilon > 0$ is a fixed constant. Thus we have

$$\operatorname*{plim}_{n\to\infty} \hat{p} = p.$$

(iii) We argue by contradiction. Assume that $\lim_{n\to\infty} \hat{p} = p$. However, no matter how large n is, there is always a positive possibility that $\hat{p} = 1$, i.e., $\mathrm{P}(\hat{p} = 1) = p^n > 0$, ($X_i = 1$ for all $i = 1, 2, \cdots, n$). Also, $\mathrm{P}(\hat{p} = 0) = (1-p)^n > 0$ for any n, ($x_i = 0$ for all $i = 1, 2, \cdots, n$).

The above results contradict the assumption that $\lim_{n\to\infty} \hat{p} = p$ for any $p \in (0, 1)$.

This example shows that we need to introduce new convergence concepts such as 'convergence in probability' in order to properly study convergence related to random variables.

Exercise 1.4.

Exercise 1.5. In the same manner used for proving (1.14), we have

$$bias\left(\hat{f}(x)\right) = h^{-1} \int f(x+hv)k(v)h dv - f(x) \tag{1.1}$$

Since $f(x)$ has a continuous second order derivative function, then by using Taylor expansion,

$$f(x+hv) = f(x) + f^{(1)}(x)hv + \frac{1}{2} f^{(2)}(\tilde{x})h^2v^2, \tag{1.2}$$

where $\tilde{x}$ is between x and $x + hv$, $f^{(2)}(x) = d^2 f(x)/dx^2$.

By (1.1), (1.2) and the dominated convergence theorem, we have

$$\begin{aligned} bias\left(\hat{f}(x)\right) &= \int f(x+hv)k(v)dv - f(x) \\ &= \int \left[f(x) + f^{(1)}(x)hv + \frac{1}{2} f^{(2)}(\tilde{x})h^2v^2\right] k(v)dv - f(x) \\ &= \frac{h^2}{2} \int f^{(2)}(\tilde{x})v^2k(v)dv \\ &= \frac{h^2}{2} f^{(2)}(x) \int v^2k(v)dv + o\left(h^2\right) \end{aligned}$$

because $\left|\int f^{(2)}(\tilde{x})v^2k(v)dv - f^{(2)}(x) \int v^2k(v)dv\right| = o(1)$ by the dominated convergence theorem.

Exercise 1.6.

Exercise 1.7.

$$\int_{-\infty}^{x}\hat{f}(v)dv = \int_{-\infty}^{x}\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{X_i-v}{h}\right)dv = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x}k\left(\frac{X_i-v}{h}\right)dv$$
$$= \frac{1}{nh}\sum_{i=1}^{n}\int_{\frac{X_i-x}{h}}^{+\infty}k(u)(-h)du = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-X_i}{h}}k(s)ds$$
$$= \frac{1}{n}\sum_{i=1}^{n}G\left(\frac{x-X_i}{h}\right)$$

where the third equality holds since we use the change of variable $\frac{X_i-v}{h}=u$, and the fourth equality holds since we use the change of variable $s=-u$ (and $k(-s)=k(s)$).

Exercise 1.8.

Exercise 1.9.

$$E\left[CV_F(h)\right] = \frac{1}{n}\sum_{i=1}^{n}\int E\left\{\left[\mathbf{1}(X_i\le x)-\hat{F}_{-i}(x)\right]^2\right\}dx,$$

where $\hat{F}_{-i}(x)=(n-1)^{-1}\sum_{j=1,j\neq i}^{n}G\left(\frac{x-X_i}{h}\right)$. Plugging this into the above equation, we get

$$\begin{aligned}
E\left[CV_F(h)\right] &= \frac{1}{n}\sum_{i=1}^{n}\int E\left\{\left[\mathbf{1}(X_i\le x)-\hat{F}_{-i}(x)\right]^2\right\}dx \\
&= \frac{1}{n(n-1)^2}\int\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i}E\left\{\left[\mathbf{1}(X_i\le x)-G\left(\frac{x-X_j}{h}\right)\right]\left[\mathbf{1}(X_i\le x)-G\left(\frac{x-X_k}{h}\right)\right]\right\}dx \\
&= \frac{1}{(n-1)}\int E\left\{\left[\mathbf{1}(X_1\le x)-G\left(\frac{x-X_2}{h}\right)\right]^2\right\}dx \\
&\quad + \frac{n-2}{(n-1)}\int E\left\{E\left\{\left[\mathbf{1}(X_1\le x)-G(\frac{x-X_2}{h})\right]\Big|X_1\right\}\right\}^2dx \\
&= CV_1+CV_2,
\end{aligned}$$

where the definitions of CV_1, CV_2 should be apparent.

Next,

$$\begin{aligned}
CV_1 &= \frac{1}{(n-1)}\int E\left\{\left[\mathbf{1}(X_1\le x)-G\left(\frac{x-X_2}{h}\right)\right]^2\right\}dx \\
&= \frac{1}{(n-1)}\int E\left\{\left[\mathbf{1}(X_1\le x)-2\mathbf{1}(X_1\le x)G\left(\frac{x-X_2}{h}\right)+\left(G\left(\frac{x-X_2}{h}\right)\right)^2\right]\right\}dx \\
&= \frac{1}{(n-1)}\int F(x)dx-\frac{2}{(n-1)}\int F(x)\left[F(x)+\frac{1}{2}\kappa_2h^2F^{(2)}(x)+o\left(h^2\right)\right]dx \\
&\quad + \frac{1}{(n-1)}\int\left[F(x)-\alpha_0hf(x)+O\left(h^2\right)\right]dx \\
&= \frac{2}{(n-1)}\int\left[F(x)-(F(x))^2\right]dx-C_1\frac{h}{n-1}+O\left(\frac{h^2}{n}\right),
\end{aligned}$$

where we have used

$$E\left[G\left(\frac{x-X_2}{h}\right)\right]=F(x)+\frac{1}{2}\kappa_2 h^2 F^{(2)}(x)+o\left(h^2\right)$$

and

$$E\left[G\left(\frac{x-X_2}{h}\right)\right]^2=F(x)-\alpha_0 h f(x)+O\left(h^2\right).$$

Finally,

$$\begin{aligned}
CV_2&=\frac{n-2}{(n-1)}\int E\left\{E\left\{\left[\mathbf{1}(X_1\le x)-G\left(\frac{x-X_2}{h}\right)\right]|X_1\right\}\right\}^2 dx\\
&=\frac{n-2}{(n-1)}\int E\left[\mathbf{1}(X_1\le x)-F(x)+\frac{1}{2}\kappa_2 h^2F^{(2)}(x)+o\left(h^2\right)\right]^2 dx\\
&=\frac{n-2}{(n-1)}\int\left[F(x)-(F(x))^2+\frac{1}{4}\kappa_2^2\left[F^{(2)}(x)\right]^2h^4+o\left(h^4\right)\right]dx.
\end{aligned}$$

Summarizing the above we have

$$\begin{aligned}
E[CV_F(h)]&=CV_1+CV_2\\
&=\frac{2}{(n-1)}\int\left[F(x)-(F(x))^2-\frac{h}{2}c_1(x)h+O\left(h^2\right)\right]dx\\
&\quad+\frac{n-2}{(n-1)}\int\left[F(x)-(F(x))^2+\frac{1}{4}\kappa_2^2\left[F^{(2)}(x)\right]^2h^4+o\left(h^4\right)\right]dx\\
&=\int F(x)(1-F(x))dx+\frac{1}{n-1}\int F(x)(1-F(x))dx-C_1hn^{-1}\\
&\quad+C_2h^4+o\left(hn^{-1}+h^4\right).
\end{aligned}$$

Exercise 1.10.

Exercise 1.11.

$$\hat{f}(x)=\frac{1}{nh_1\cdots h_q}\sum_{i=1}^n k\left(\frac{X_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{X_{iq}-x_q}{h_1}\right).$$

$$\begin{aligned}
bias\left(\hat{f}(x)\right)&=E\left(\hat{f}(x)\right)-f(x)\\
&=E\left[\frac{1}{nh_1\cdots h_q}k\left(\frac{X_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{X_{iq}-x_q}{h_q}\right)\right]-f(x)\\
&=\int f(x_i)\frac{1}{nh_1\cdots h_q}k\left(\frac{x_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{x_{iq}-x_q}{h_q}\right)dx_i-f(x)\\
&=\int f(x+hv)k(v)dv-f(x)\\
&=\int\left[f(x)+\sum_{s=1}^q f_s(x)h_sv_s+\frac{1}{2}\sum_{s=1}^q\sum_{t=1}^q f_{st}(x)h_sh_tv_sv_t\right]k(v)dv-f(x)+O\left(\sum_{s=1}^q h_s^3\right)\\
&=\frac{\kappa_2}{2}\sum_{s=1}^q f_{ss}(x)h_s^2+O\left(\sum_{s=1}^q h_s^3\right),
\end{aligned}$$

where $f_{st}(x) = \partial^2 f(x)/\partial x_s \partial x_t$, the third equality holds since we use the change of variable $\frac{X_{is}-x_s}{h_s} = v_s$, the fifth equality holds since

$$f(x+hv) = f(x) + \sum_{s=1}^{q} f_s(x) h_s v_s + \frac{1}{2}\sum_{s=1}^{q}\sum_{t=1}^{q} f_{st}(x) h_s h_t v_s v_t. + O\left(\sum_{s=1}^{q} h_s^3\right).$$

$$\begin{aligned}
Var\left(\hat{f}(x)\right) &= Var\left[\frac{1}{nh_1\cdots h_q}\sum_{k=1}^{n} k\left(\frac{X_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{X_{iq}-x_q}{h_q}\right)\right] \\
&= \frac{1}{nh_1^2\cdots h_q^2} Var\left[k\left(\frac{X_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{X_{iq}-x_q}{h_q}\right)\right] \\
&= \frac{1}{nh_1^2\cdots h_q^2} E\left[k\left(\frac{X_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{X_{iq}-x_q}{h_q}\right)\right]^2 \\
&\quad - \frac{1}{nh_1^2\cdots h_q^2}\left\{E\left[k\left(\frac{X_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{X_{iq}-x_q}{h_q}\right)\right]\right\}^2 \\
&= \frac{1}{nh_1^2\cdots h_q^2}\int f(x_i)\left[k\left(\frac{x_{i1}-x_1}{h_1}\right)\times\cdots\times k\left(\frac{x_{iq}-x_q}{h_q}\right)\right]^2 dx_i + o\left(\frac{1}{nh_1\cdots h_q}\right) \\
&= \frac{\kappa^q}{nh_1\cdots h_q} f(x) + o\left(\sum_{s=1}^{q} h_s^2 + \frac{1}{nh_1\cdots h_q}\right).
\end{aligned}$$

Exercise 1.12.

Exercise 1.13. We assume that f is continuous on $[0,1]$ and that $f(0) \neq 0$. We know that $\hat{f}(x) = (nh)^{-1}\sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)$. Hence,

$$\hat{f}(0) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{X_i}{h}\right).$$

$$\begin{aligned}
E[\hat{f}(0)] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{X_i}{h}\right)\right] = E\left[\frac{1}{h}k\left(\frac{X_1}{h}\right)\right] \\
&= \int_0^1 \frac{1}{h}k\left(\frac{x_1}{h}\right) f(x_1)dx_1 = \int_0^{1/h}\frac{1}{h}k(u)f(hu)hdu \\
&= \int_0^{1/h} k(u)f(hu)du = \int_0^{1/h} k(u)f(0)du + h\int_0^{1/h} f^{(1)}(\tilde{x})uk(u)du \\
&\to f(0)\int_0^{\infty} k(u)du = \frac{1}{2}f(0) \neq f(0) \text{ as } h \to 0.
\end{aligned}$$

Therefore, $\hat{f}(0)$ is a biased (even asymptotically) and inconsistent estimator for $f(0)$. We need to use a boundary kernel to get a consistent estimator for $f(0)$ as the next exercise shows.

Exercise 1.14.

Exercise 1.15.

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

where $k(\cdot)$ is a νth order kernel function.

$$\begin{aligned}
bias\left[\hat{f}(x)\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{X_i - x}{h}\right)\right] - f(x) = E\left[\frac{1}{h}k\left(\frac{X_1 - x}{h}\right)\right] - f(x)\\
&= \int \frac{1}{h}k\left(\frac{x_1 - x}{h}\right) f(x_1)dx_1 - f(x) = \int k(u)f(x+hu)du - f(x)\\
&= \int k(u)\left[f(x) + f^{(1)}(x)hu + \cdots + \frac{1}{\nu!}f^{(\nu)}(x)h^\nu u^\nu + o(h^\nu)\right]du - f(x)\\
&= f(x) + \frac{1}{\nu!}f^{(\nu)}(x)h^\nu \int k(u)u^\nu du + o(h^\nu) - f(x)\\
&= \frac{1}{\nu!}f^{(\nu)}(x)h^\nu \int k(u)u^\nu du + o(h^\nu)\\
&= O(h^\nu).
\end{aligned}$$

$$\begin{aligned}
Var\left[\hat{f}(x)\right] &= Var\left[\frac{1}{n}\sum_{s=1}^{q}\frac{1}{h}k\left(\frac{X_i - x}{h}\right)\right] = \frac{1}{n^2}\cdot nVar\left[\frac{1}{h}k\left(\frac{X_i - x}{h}\right)\right]\\
&= \frac{1}{n}\left\{E\left[\frac{1}{h}k\left(\frac{X_i - x}{h}\right)\right]^2 - \left(E\left[\frac{1}{h}k\left(\frac{X_i - x}{h}\right)\right]\right)^2\right\}\\
&= \frac{1}{n}\left\{\int\left[\frac{1}{h}\left(\frac{x_i - x}{h}\right)\right]^2 f(x_i)dx_i - \left(\int\left[\frac{1}{h}k\left(\frac{x_i - x}{h}\right)\right]f(x_i)dx_i\right)^2\right\}\\
&= \frac{1}{n}\left\{\int\left[\frac{1}{h}k(u)\right]^2 f(x+hu)hdu - \left(\int\left[\frac{1}{h}k(u)\right]f(x+hu)hdu\right)^2\right\}\\
&= \frac{1}{nh}\int [k(u)]^2 f(x+hu)du + O\left(h^2(nh)^1\right)\\
&= \frac{\kappa}{nh}f(x) + o\left((nh)^{-1}\right)\\
&= O\left(\frac{1}{nh}\right),
\end{aligned}$$

where $\kappa = \int k(v)^2 dv$.

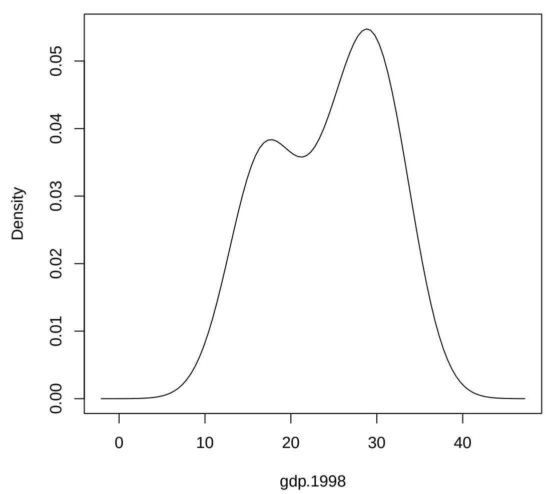
Exercise 1.16.

Exercise 1.17. Consider the Italy data from Section 1.13.5 (included in the np package) and the rule-of-thumb bandwidth selector for GDP for the year 1998 (i.e., the last year in the panel) using the default second-order Gaussian kernel:

```
R> data(Italy)
R> gdp.1998 <- subset(Italy, year==1998)$gdp
R> bw.rt <- npudensbw(~gdp.1998,bwmethod="normal-reference")
R> summary(bw.rt)

Data (21 observations, 1 variable(s)):

Bandwidth Selection Method: Normal Reference
Formula: ~gdp.1998
Bandwidth Type: Fixed

Var. Name: gdp.1998 Bandwidth: 3.63 Scale Factor: 1.06

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Vars.: 1
```

We see that the normal-reference rule-of-thumb for gdp is h=3.63. We can plot the estimate as follows:[1]

```
R> fhat <- npudens(~gdp.1998,bws=bw.rt)
R> plot(fhat,neval=100,xtrim=-0.75)
```

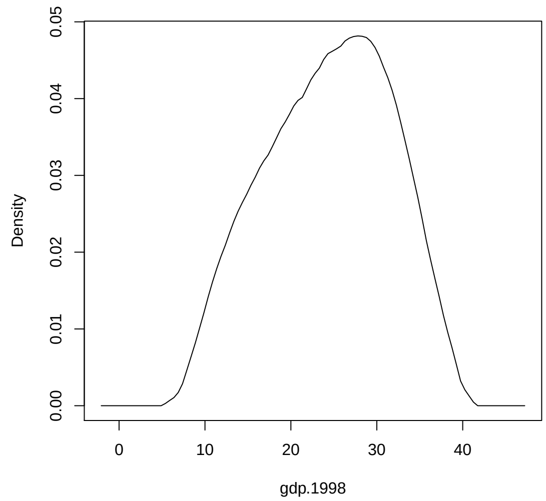


Italy GDP data (1998), h=3.63

Now you can try larger manual bandwidths. Trial and error indicates that a bandwidth of approximately 4.1 (i.e., greater than 1.13 times the normal-reference rule-of-thumb bandwidth) appears to obscure the bimodal nature of the relationship as the following plot demonstrates:

[1] Note that the use of `xtrim` alters the domain for the density estimate by extending it by the data's 75th percentile in either direction so that the density estimate touches the horizontal axis. Also, the use of `neval` increases the number of evaluation points above the default resulting in a slightly smoother estimate in between the two modes.

```
R> fhat <- npudens(~gdp.1998,bws=4.1)
R> plot(fhat,neval=100,xtrim=-0.75)
```



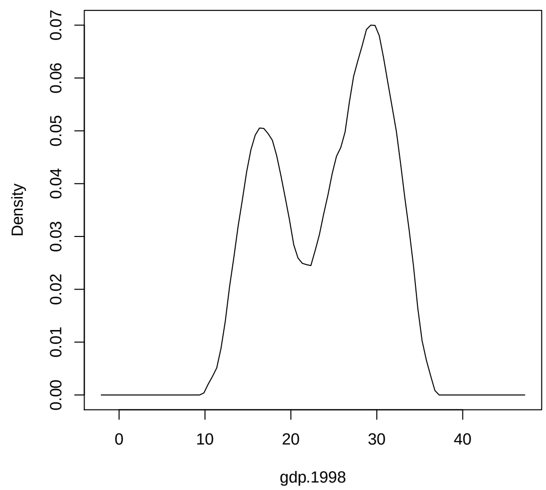
Italy GDP data (1998), $h = 4.1$

Finally, we can plot the estimate based on least squares cross-validation. First we obtain the bandwidth object:

```
R> bw.cv <- npudensbw(~gdp.1998,bwmethod="cv.ls")
R> summary(bw.cv)

Data (21 observations, 1 variable(s)):

Bandwidth Selection Method: Least Squares Cross-Validation
Formula: ~gdp.1998
Bandwidth Type: Fixed
Objective Function Value: 0.0446 (achieved on multistart 1)

Var. Name: gdp.1998 Bandwidth: 1.99 Scale Factor: 0.58

Continuous Kernel Type: Second-Order Epanechnikov
No. Continuous Vars.: 1
```

Note that the cross-validated bandwidth h=1.99 is smaller than that for the normal reference rule-of-thumb (i.e., 0.55 times the rule-of-thumb bandwidth), and the resulting plot displays a more pronounced bimodal structure:

```
R> fhat <- npudens(~gdp.1998,bws=bw.cv)
R> plot(fhat,neval=100,xtrim=-0.75)
```



Italy GDP data (1998), h=1.99

Judging from the above plot and a histogram of the same, the cross-validated estimate appears to be sensible.

Chapter 2

Regression: Solutions

Exercise 2.1. By definition we have $\hat{m}_1(x) = (nh_1 \cdots h_q)^{-1} \sum_{i=1}^n (g(X_i) - g(x))K\left(\frac{X_i - x}{h}\right)$, and by (2.8) and (2.9), we have

$$E\left[\hat{m}_1(x)\right] = f(x)\sum_{s=1}^q h_s^2 B_s(x) + O\left(\sum_{s=1}^q h_s^3\right) \quad \text{and} \quad Var[\hat{m}_1(x)] = O\left((nh_1 \cdots h_q)^{-1}\sum_{s=1}^q h_s^2\right).$$

Hence,

$$\begin{aligned}
E\left[\hat{m}_1(x) - f(x)\sum_{s=1}^q h_s^2 B_s(x)\right]^2 &= \left(E\left[\hat{m}_1(x) - f(x)\sum_{s=1}^q h_s^2 B_s(x)\right]\right)^2 + Var\left[\hat{m}_1(x)\right] \\
&= O\left(\left(\sum_{s=1}^q h_s^3\right)^2\right) + O\left((nh_1 \cdots h_q)^{-1}\sum_{s=1}^q h_s^2\right) \\
&= O\left(\left(\sum_{s=1}^q h_s^3\right)^2 + (nh_1 \cdots h_q)^{-1}\sum_{s=1}^q h_s^2\right) \\
&= O\left(\eta_2^3 + \eta_1\eta_2\right). \qquad (2.1)
\end{aligned}$$

By Lemma A.7 (ii) (Li and Racine (2007, page 686)) we know that (2.1) implies that $\hat{m}_1(x) - f(x)\sum_{s=1}^q h_s^2 B_s(x) = O_p\left(\eta_2^{3/2} + \eta_1^{1/2}\eta_2^{1/2}\right)$. Or equivalently,

$$\hat{m}_1(x) = f(x)\sum_{s=1}^q h_s^2 B_s(x) + O_p\left(\eta_2^{3/2} + \eta_1^{1/2}\eta_2^{1/2}\right).$$

Exercise 2.2.

Exercise 2.3. Let $K_{h,ij} = K_{h,x_i x_j} = \prod_{s=1}^{q} h_s^{-1} k((X_{is} - X_{js})/h_s)$, we have

$$
\begin{aligned}
& n^{-1} \sum_{i=1}^{n} u_i \left(g(X_i) - \hat{g}_{-i}(X_i)\right) M(X_i) \\
&\quad = \frac{1}{(n-1)n} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} u_i \left(g(X_i) - g(X_j)\right) K_{h,ij} M(X_i)/f(X_i) + (s.o.) \\
&\quad = \frac{1}{(n-1)n} \sum_{i} \sum_{j\neq i} u_i (g_i - g_j) K_{h,ij} M(X_i)/f(X_i) \\
&\qquad - \frac{1}{(n-1)n} \sum_{i} \sum_{j\neq i} u_i u_j K_{h,ij} M(X_i)/f(X_i) + (s.o.) \\
&\quad = I_1 - I_2 + (s.o.),
\end{aligned}
$$

where $I_1 = ((n-1)n)^{-1} \sum_i \sum_{j\neq i} u_i (g_i - g_j) K_{h,ij} M(X_i)/f(X_i)$ and (s.o.) denotes smaller order terms.

$$
\begin{aligned}
E\left(I_1^2\right) &= \frac{1}{(n-1)^2 n^2} E\left[\sum_i \sum_{j\neq i} \sum_{l\neq i} u_i^2 (g_i - g_j) K_{h,ij} (g_i - g_l) K_{h,il} [M(X_i)]^2/[f(X_i)]^2\right] \\
&= \frac{1}{(n-1)^2 n^2} \sum_i E\left[\sum_{j\neq i} \sum_{l\neq i, l\neq j} \sigma^2(X_i)(g_i - g_j) K_{h,ij} (g_i - g_l) K_{h,il} [M(X_i)]^2/[f(X_i)]^2\right] \\
&\quad + \frac{1}{(n-1)^2 n^2} \sum_i E\left[\sum_{j\neq i} \sigma^2(X_i)(g_i - g_j)^2 K_{h,ij}^2 [M(X_i)]^2/[f(X_i)]^2\right] \\
&= \frac{1}{(n-1)^2 n^2} \sum_i E\left[\sum_{j\neq i} \sum_{l\neq i, l\neq j} \sigma^2(X_i) \left(E((g_i - g_j) K_{h,ij})\right)^2 [M(X_i)]^2/[f(X_i)]^2 | X_i\right] \\
&\quad + \frac{1}{(n-1)^2 n} n(n-1) O\left((h_1 \cdots h_q)^{-1}\right) + (s.o.) \\
&= \frac{1}{n} O\left(\left(\sum_{s=1}^{q} h_s^2\right)^2\right) + O\left(\left(n^2 h_1 \cdots h_q\right)^{-1}\right) + (s.o.) \\
&= O\left(n^{-1}\eta_2^2 + n^{-1}\eta_1\right).
\end{aligned}
$$

The above result implies that

$$
I_1 = O_p\left(n^{-1/2}\eta_2 + n^{-1/2}\eta_1^{1/2}\right).
$$

Next,

$$E\left(I_2^2\right) = \frac{1}{(n-1)^2n^2}\sum_i\sum_{j\neq i}E\left[u_i^2u_j^2K_{h,ij}^2M_i^2/f_i^2\right] + \frac{1}{(n-1)^2n^2}\sum_i\sum_{j\neq i}E\left[u_i^2u_j^2K_{h,ij}^2M_iM_j/f_if_j\right]$$
$$= \frac{1}{(n-1)^2n^2}\sum_i\sum_{j\neq i}E\left[\sigma^2(X_i)\sigma^2(X_j)K_{h,ij}^2M_i^2/f_i^2\right]$$
$$+ \frac{1}{(n-1)^2n^2}\sum_i\sum_{j\neq i}E\left[\sigma^2(X_i)\sigma^2(X_j)K_{h,ij}^2M_iM_j/f_if_j\right]$$
$$= O\left(\left(n^2h_1\cdots h_q\right)^{-1}\right).$$

Hence,

$$I_2 = O_p\left(n^{-1/2}\eta_1^{1/2}\right).$$

Summarizing the above we have shown that

$$n^{-1}\sum_{i=1}^n u_i\left(g(X_i)-\hat{g}_{-i}(X_i)\right)M(X_i) = I_1 - I_2 + (s.o.) = O_p\left(n^{-1/2}\eta_2 + n^{-1/2}\eta_1^{1/2}\right).$$

Exercise 2.4.

Exercise 2.5.

(i) $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^n \bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}$. By the independence of $\bar{X}_i$ and $\tilde{X}_i$, we have

$$E\left[\hat{f}(x)\right] = E\left[\frac{1}{n}\sum_{i=1}^n \bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}\right] = E\left[\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\ \ ,x_i}\right] = E\left[\bar{K}_{h,x\,x_i}\right]E\left[\tilde{K}_{h,x\,x_i}\right]$$
$$= \bar{f}(\bar{x})E\left[\tilde{K}_{h,x\,x_i}\right] + O\left(\bar{h}^2\right).$$

(ii) $\hat{m}_1(x) = \frac{1}{n}\sum_{i=1}^n\left(\bar{g}_i - \bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}$.

$$E\left[\hat{m}_1(x)\right] = E\left[\frac{1}{n}\sum_{i=1}^n\left(\bar{g}_i-\bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}\right]$$
$$= E\left[\left(\bar{g}_i-\bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}\right]$$
$$= E\left[\left(\bar{g}_i-\bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\right]E\left[\tilde{K}_{h,x\,x_i}\right]$$
$$= \sum_{s=1}^{q_1}\bar{B}_s(\bar{x})h_s^2\bar{f}(\bar{x})E\left[\tilde{K}_{h,x\,x_i}\right] + o\left(\eta_{2,q_1}\right).$$

$$
\begin{aligned}
E\left[\hat{m}_1(x)^2\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\bar{g}_i - \bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}\left(\bar{g}_j - \bar{g}(\bar{x})\right)\bar{K}_{h,xj}\tilde{K}_{h,xj}\right] \\
&= E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\left(\bar{g}_i - \bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}\left(\bar{g}_j - \bar{g}(\bar{x})\right)\bar{K}_{h,xj}\tilde{K}_{h,xj}\right] \\
&\quad + E\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(\bar{g}_i - \bar{g}(\bar{x})\right)^2\bar{K}^2_{h,x\,x_i}\tilde{K}^2_{h,x\,x_i}\right] \\
&= \frac{n(n-1)}{n^2}\left\{E\left[\left(\bar{g}_i - \bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}\right]\right\}^2 \\
&\quad + \frac{1}{n}E\left[(\bar{g}_i - \bar{g}(\bar{x}))^2\bar{K}^2_{h,x\,x_i}\tilde{K}^2_{h,x\,x_i}\right] \\
&= \frac{n(n-1)}{n^2}\left\{E\left[\left(\bar{g}_i - \bar{g}(\bar{x})\right)\bar{K}_{h,x\,x_i}\right]\right\}^2\left\{E\left[\tilde{K}_{h,x\,x_i}\right]\right\}^2 \\
&\quad + \frac{1}{n}E\left[\left(\bar{g}_i - \bar{g}(\bar{x})\right)^2\bar{K}^2_{h,x\,x_i}\right]E\left[\tilde{K}^2_{h,x\,x_i}\right] \\
&= \left[\sum_{s=1}^{q_1}\bar{B}_s(\bar{x})h_s^2\bar{f}(\bar{x})\right]^2\left\{E\left[\tilde{K}_{h,x\,x_i}\right]\right\}^2 + o\left(\eta^2_{2,q_1}\right) \\
&\quad + \frac{E\left[\tilde{K}^2_{h,x\,x_i}\right]\left(\int v^2k^2(v)dv\right)}{nh_1\cdots h_{q_1}}\sum_{s=1}^{q_1}\bar{g}^2_s(\bar{x})\bar{f}(\bar{x})h_s^2 + o\left(\eta^2_{2,q_1}\eta_{1,q_1}\right).
\end{aligned}
$$

(iii) $\hat{m}_2(x) = \frac{1}{n}\sum_{i=1}^{n} u_i\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}$.

$$
\begin{aligned}
E\left[\hat{m}_2(x)^2\right] &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}u_i\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}\right)^2\right] \\
&= E\left[\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}u_i\bar{K}_{h,x\,x_i}\tilde{K}_{h,x\,x_i}u_j\bar{K}_{h,xj}\tilde{K}_{h,xj}\right)\right] \\
&= E\left[\left(\frac{1}{n^2}\sum_{i=1}^{n}u_i^2\bar{K}^2_{h,x\,x_i}\tilde{K}^2_{h,x\,x_i}\right)\right] \\
&= \frac{1}{n}E\left[\left(u_i^2\bar{K}^2_{h,x\,x_i}\right)\right]E\left(\tilde{K}^2_{h,x\,x_i}\right) \\
&= \eta_{1,q_1}\kappa^{q_1}\bar{\sigma}^2(\bar{x})\bar{f}(\bar{x})E\left(\tilde{K}^2_{h,x\,x_i}\right)[1 + o(1)].
\end{aligned}
$$

(iv)

$$
\begin{aligned}
E\left[A_n(x)\right]^2 &= E\left[\hat{m}_1(x)+\hat{m}_2(x)\right]^2 / \left[\bar{f}(\bar{x})E(\tilde{K}_{h,x\,x_i})\right]^2 \\
&= E\left[\hat{m}_1(x)^2+2\hat{m}_1(x)\hat{m}_2(x)+\hat{m}_2(x)^2\right] / \left[\bar{f}(\bar{x})E(\tilde{K}_{h,x\,x_i})\right]^2 \\
&= \left[E(\hat{m}_1(x)^2)+0+E(\hat{m}_2(x)^2)\right] / \left[\bar{f}(\bar{x})E(\tilde{K}_{h,x\,x_i})\right]^2 \\
&= \left[\sum_{s=1}^{q_1}\bar{B}_s(\bar{x})h_s^2\right]^2 + \frac{\left(\int v^2k^2(v)dv\right)}{nh_1\cdots h_{q_1}\bar{f}(\bar{x})}\left(\sum_{s=1}^{q_1}\bar{g}_s^2(\bar{x})h_s^2\right)\frac{E\left[\tilde{K}_{h,x\,x_i}^2\right]}{\left[E\left(\tilde{K}_{h,x\,x_i}\right)\right]^2} \\
&\quad + \frac{\kappa^{q_1}\bar{\sigma}^2(\bar{x})}{nh_1\cdots h_{q_1}\bar{f}(\bar{x})}\frac{E\left[\tilde{K}_{h,x\,x_i}^2\right]}{\left[E\left(\tilde{K}_{h,x\,x_i}\right)\right]^2}+(s.o.) \\
&= \left[\sum_{s=1}^{q_1}\bar{B}_s(\bar{x})h_s^2\right]^2 + \frac{\left(\int v^2k^2(v)dv\right)\left(\sum_{s=1}^{q_1}\bar{g}_s^2(\bar{x})h_s^2\right)+\kappa^{q_1}\bar{\sigma}^2(\bar{x})}{nh_1\cdots h_{q_1}\bar{f}(\bar{x})}\frac{E\left[\tilde{K}_{h,x\,x_i}^2\right]}{\left[E\left(\tilde{K}_{h,x\,x_i}\right)\right]^2}+(s.o.).
\end{aligned}
$$

(v) From (2.112), we know that in order to minimize CV, $h_{q_1+1},\cdots,h_q$ must minimize $R(\tilde{x},h_{q_1+1},\cdots,h_q)=\frac{E\left[\tilde{K}_{h,x\,x_i}^2\right]}{\left[E\left(\tilde{K}_{h,x\,x_i}\right)\right]^2}$. It is obvious that $E\left[\tilde{K}_{h,x\,x_i}^2\right]/\left[E\left(\tilde{K}_{h,x\,x_i}\right)\right]^2\geq 1$, and when $(h_{q_1+1},\cdots,h_q)=(\infty,\cdots,\infty)$, $E\left[\tilde{K}_{h,x\,x_i}^2\right]/\left[E\left(\tilde{K}_{h,x\,x_i}\right)\right]^2=1$. Define $Z_n=\tilde{K}_{h,x\,x_i}$. If one of the h_s (for $s=q_1+1,\ldots,q$) does not goes to ∞, then it is easy to show that $Var(Z_n)=E\left[Z_n^2\right]=\left[E\left(Z_n\right)\right]^2>0$, which is equivalent to that $E\left[\tilde{K}_{h,x\,x_i}^2\right]/\left[E\left(\tilde{K}_{h,x\,x_i}\right)\right]^2>1$.

Thus, $(h_{q_1+1},\cdots,h_q)=(\infty,\cdots,\infty)$ is the unique solution.

Exercise 2.6.

Exercise 2.7.

(i) $A_{11}^{1,x}=n^{-1}\sum_i K_{h,ix}\equiv\hat{f}(x)$. Hence, we know that $A_{11}^{1,x}=\hat{f}(x)=f(x)+o_p(1)$.

(ii) $A_{12}^{1,x}=n^{-1}\sum_i K_{h,ix}(X_i-x)'$. Hence,

$$
\begin{aligned}
E\left[A_{12}^{1,x}\right] &= n^{-1}\sum_i E\left[K_{h,ix}(X_i-x)'\right] \\
&= E\left[K_{h,ix}(X_i-x)'\right] \\
&= (h_1\ldots h_q)^{-1}\int f(x_i)K\left(\frac{x_i-x}{h}\right)(x_i-x)'dx_i \\
&= \int f(x+hv)K(v)(hv)'dv \\
&= 0+\kappa_2\left(vec\left(h_s^2f_s(x)\right)\right)'+O\left(|h|^3\right),
\end{aligned}
$$

where $vec(h_s^2 f_s(x))$ is a $q \times 1$ vector with the s^{th} position given by $h_s^2 f_s(x)$, $\kappa_2 = \int k(v_s) v_s^2 dv_s$ and $|h| = \sum_{s=1}^q h_s$.

It is easy to show that $Var(A_{12}^{1,x}) = O\left(|h|^2 (nh_1 \dots h_q)\right)$. Hence,

$$A_{12}^{1,x} = O_p\left(|h|^2 + |h|(nh_1 \dots h_q)^{-1/2}\right) = O_p\left(\eta_2 + \eta_2^{1/2}\eta_1^{1/2}\right).$$

(iii) By noting that D_h^{-2} is a $q \times q$ diagonal matrix with the s^{th}-diagonal element given by h_s^{-2}, then using the same proof as in (ii) above, one can show that

$$E(A_{21}^{1,x}) = \kappa_2 vec(f_s(x)) + O(|h|) \equiv \kappa_2 f^{(1)}(x) + O(|h|).$$

Also, it is easy to show that $Var(A_{21}^{1,x}) = O\left(\left(n\left(h_1 \dots h_q\right)|h|^2\right)^{-1}\right) = o(1)$ provided that $n(h_1 \dots h_q)|h|^2 \to \infty$ as $n \to \infty$. Hence,

$$A_{21}^{1,x} = \kappa_2 f^{(1)}(x) + o_p(1).$$

(iv) It is easy to see that

$$\begin{aligned}
E(A_{22}^{1,x}) &= E\left[K_{h,ix} D_h^{-2}(X_i - x)(X_i - x)'\right] \\
&= (h_1 \dots h_q)^{-1} \int K(v) D_h^{-2}(hv)(hv)'(h_1 \dots h_q) dv \\
&= \kappa_2 f(x) I_q \left(\int k(v_s) v_s^2 dv_s f(x)\right) + O\left(|h|^2\right) \\
&\equiv \kappa_2 f(x) + o(1),
\end{aligned}$$

where $Diag(b_s)$ denotes a diagonal matrix with the s^{th} diagonal element equals to b_s.

Noting that $A_{22}^{1,x}$ is a $q \times q$ matrix, it is easy to show that for each component of $A_{22}^{1,x}$ (which is a scalar), its variance is of the order of $O\left((nh_1 \dots h_q)^{-1}\right)$. Hence, we know that

$$A_{22}^{1,x} = \kappa_2 f(x) I_q + O_p\left(|h|^2 + (nh_1 \dots h_q)^{-1/2}\right).$$

Exercise 2.8.

Exercise 2.9.

(i)

$$\begin{aligned}
Var\left(A_1^{3,x}\right) &= \left(nh_1^2 \dots h_q^2\right)^{-1} E\left[K_{h,ix}^2 u_i^2\right] \\
&= \left(nh_1^2 \dots h_q^2\right)^{-1} \int f(x_i) K\left((x_i - x)/h\right)^2 \sigma^2(x_i) dx_i \\
&= \left(nh_1^2 \dots h_q^2\right)^{-1} \int f(x + hv) K(v)^2 \sigma^2(x + hv)(h_1 \dots h_q) dv \\
&= (nh_1 \dots h_q)^{-1} \left[f(x)\sigma^2(x) \prod_{s=1}^q \int k(v_s)^2 dv_s + O(|h|)\right] \\
&= (nh_1 \dots h_q)^{-1} \kappa^q f(x)\sigma^2(x) + o\left((nh_1 \dots h_q)^{-1}\right),
\end{aligned}$$

which is equivalent to

$$Var\left((nh_1 \dots h_q)^{1/2} A_1^{3,x}\right) = \kappa^q f(x)\sigma^2(x) + o(1).$$

(ii)

$$\begin{aligned}
Var\left(D_h A_2^{3,x}\right) &= \left(nh_1^2 \dots h_q^2\right)^{-1} E\left[K_{h,ix}^2 D_h^{-1}(X_i - x)(X_i - x)' D_h^{-1} u_i^2\right] \\
&= \left(nh_1^2 \dots h_q^2\right)^{-1} \int f(x_i) K\left((x_i - x)/h\right)^2 \sigma^2(x_i) D_h^{-1}(x_i - x)(x_i - x)' D_h^{-1} dx_i \\
&= \left(nh_1^2 \dots h_q^2\right)^{-1} \int f(x + hv) K(v)^2 \sigma^2(x + hv) D_h^{-1}(hv)(hv)' D_h^{-1}(h_1 \dots h_q) dv \\
&= (nh_1 \dots h_q)^{-1} \left\{ Diag[f(x)\sigma^2(x) \int k(v_s)^2 v_s^2 dv_s \prod_{j \neq s} \int k(v_j)^2 dv_j] + O(|h|) \right\} \\
&= (nh_1 \dots h_q)^{-1} \kappa^{q-1} \kappa_{22} f(x)\sigma^2(x) + o\left((nh_1 \dots h_q)^{-1}\right),
\end{aligned}$$

where $\kappa_{22} = \int k(v_s)^2 v_s^2 dv_s$. The above result is equivalent to

$$Var\left((nh_1 \dots h_q)^{1/2} D_h A_2^{3,x}\right) = \kappa^q f(x)\sigma^2(x) + o(1).$$

(iii)

$$\begin{aligned}
Cov\left(A_1^{3,x}, D_h A_2^{3,x}\right) &= \left(nh_1^2 \dots h_q^2\right)^{-1} E\left[K_{h,ix}^2 D_h^{-1}(X_i - x)' u_i^2\right] \\
&= \left(nh_1^2 \dots h_q^2\right)^{-1} \int f(x_i) K\left((x_i - x)/h\right)^2 \sigma^2(x_i) D_h^{-1}(x_i - x) dx_i \\
&= \left(nh_1^2 \dots h_q^2\right)^{-1} \int f(x + hv) K(v)^2 \sigma^2(x + hv) D_h^{-1}(hv)(h_1 \dots h_q) dv \\
&= (nh_1 \dots h_q)^{-1} \{0 + o(|h|)\} \\
&= O\left(|h|(nh_1 \dots h_q)^{-1}\right),
\end{aligned}$$

where in the fourth equality above we used $\int K^2(v) v_s dv = 0$ because $K(v)^2$ is an even function (since $K(v)$ is even). The above result is equivalent to

$$Cov\left((nh_1 \dots h_q)^{1/2} A_1^{3,x}, (nh_1 \dots h_q)^{1/2} D_h A_2^{3,x}\right) = o(1).$$

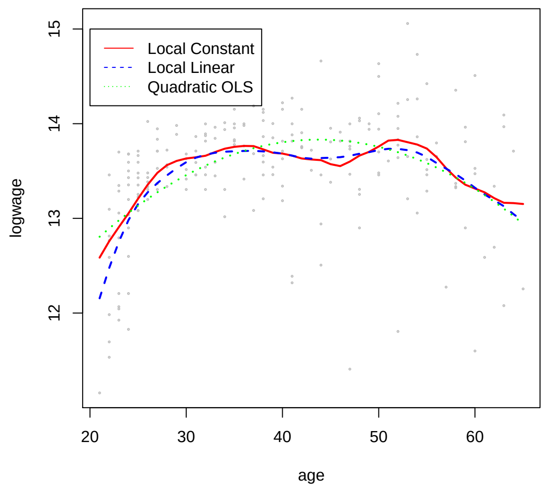
Exercise 2.10.

Exercise 2.11.

(i) Compute and plot the local constant, local linear, and parametric quadratic estimates using least squares cross-validated bandwidths for the kernel estimators.

```
R> data(cps71)
R> attach(cps71)
R> model.lc <- npreg(logwage~age,regtype="lc")
R> model.ll <- npreg(logwage~age,regtype="ll")
R> model.ols <- lm(logwage~age+I(age^2))
```

```
R> plot(age,logwage,cex=0.25,col="gray")
R> lines(age,fitted(model.lc),lty=1,lwd=2,col="red")
R> lines(age,fitted(model.ll),lty=2,lwd=2,col="blue")
R> lines(age,fitted(model.ols),lty=3,lwd=2,col="green")
R> legend(20,15,c("Local Constant","Local Linear","Quadratic OLS"),
+         lty=c(1,2,3),
+         col=c("red","blue","green"))
```

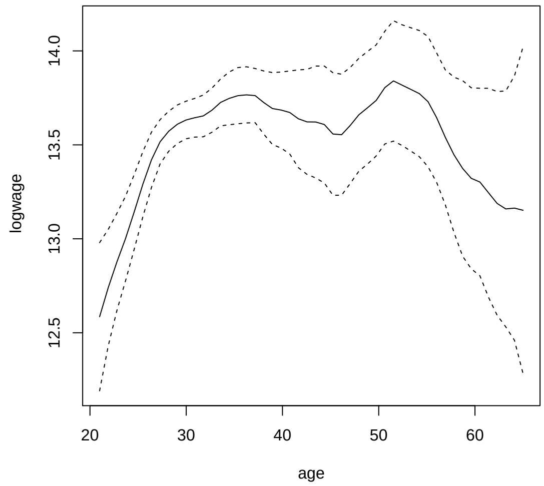


(ii) Is the dip present in the resulting nonparametric estimates?

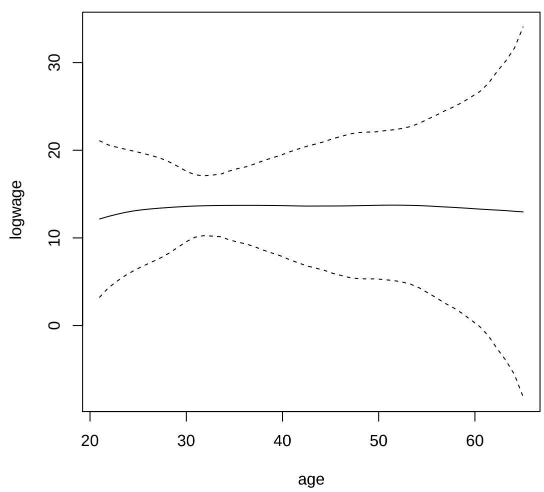
Yes, there is a dip present in the nonparametric estimates.

(iii) Plot the nonparametric estimates along with their error bounds using the asymptotic formulas for the standard errors (i.e., $\hat{g}(x) \pm 2\hat{\sigma}(\hat{g}(x))$). Without conducting a formal test, does the dip appear to be significant?

```
R>      plot(model.lc,plot.errors.method="asymptotic",
+           plot.errors.style="band")
```




```
R>      plot(model.ll,plot.errors.method="asymptotic",
+           plot.errors.style="band")
```



It does not appear, based on the pointwise standard error bounds, that the dip is significant (i.e., a line segment connecting the top of the two 'humps' would lie within the standard error bars).

(iv) Which nonparametric estimator (i.e., the local constant or local linear) appears to provide the most "appropriate" fit to this data?

The local linear estimator appears to be doing better fitting the data for low values of age, while it is also less noisy (i.e., smoother).

Chapter 3

Frequency Estimation with Mixed Data: Solutions

Exercise 3.1. The hint in Li and Racine (2007, page 124) is a solution to this problem.

Exercise 3.2.

Chapter 4

Kernel Estimation with Mixed Data: Solutions

Exercise 4.1. $E(\tilde{p}(x)) = E\left[\mathbf{1}(X_i = x)\right] = \sum_{x_i \in \mathcal{S}} p(x_i)\mathbf{1}(x_i = x) = p(x)$ and $Var(\tilde{p}(x)) = n^{-1}Var(\mathbf{1}(X_i = x)) = n^{-1}\left\{E\left[\mathbf{1}(X_i = x)\right] - \left[E(\mathbf{1}(X_i = x))\right]^2\right\} = n^{-1}\left[p - p^2\right] = n^{-1}p(1-p)$. Hence, $MSE\left(\tilde{p}(x)\right) = n^{-1}p(1-p) = O\left(n^{-1}\right)$.

Exercise 4.2.

Exercise 4.3.

(i) A short proof is given by

$E\left[\hat{p}(x)\right] = E\left[n^{-1}\sum_{i=1}^{n} L\left(X_i, x, \lambda\right)\right] = E\left[L\left(X_i, x, \lambda\right)\right] = \sum_{x_i \in \mathcal{S}} p(x_i)L\left(X_i, x, \lambda\right) = p(x)\prod_{s=1}^{r}(1 - \lambda_s) + \sum_{s=1}^{r} \lambda_s \prod_{t \neq s} \frac{1-\lambda_t}{c_t - 1} \sum_{z^d} \mathbf{1}_s(x^d, z^d) + O\left(|\lambda|^2\right) = p\left(x^d\right) + \lambda_s \sum_{s=1}^{r} B_{p,s}$, where $B_{p,s} = \sum_{z^d} \frac{1}{c_s - 1}\mathbf{1}_s(x^d, z^d) - p\left(x^d\right)$.

We also give a longer proof below. From (4.3) $\hat{p}\left(x^d\right)=\frac{1}{n}\sum\limits_{i=1}^{n}L\left(X_i^d,x^d,\lambda\right)$, we have

$$
\begin{aligned}
E\left(\hat{p}\left(x^d\right)\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}L\left(X_i^d,x^d,\lambda\right)\right)=E\left(L\left(X_1^d,x^d,\lambda\right)\right)\\
&= \sum_{z^d\in S^d}\Big\{\prod_{t=1}^{r}\mathbf{1}(x_t^d=z_t^d)\prod_{t=1}^{r}(1-\lambda_t)+\sum_{s=1}^{r}\mathbf{1}(x_s^d\neq z_s^d)\prod_{t\neq s}^{r}\mathbf{1}(x_t=z_t^d)\frac{\lambda_s}{c_s-1}\prod_{t\neq s}^{r}(1-\lambda_t)\\
&\quad+\sum_{1\le i_1<i_2\le r}\mathbf{1}(x_{i_1}^d\neq z_{i_1}^d)\mathbf{1}(x_{i_2}^d\neq z_{i_2}^d)\prod_{t\neq i_1,i_2}\mathbf{1}(x_t^d=z_t^d)\frac{\lambda_{i_1}}{c_{i_1}-1}\frac{\lambda_{i_2}}{c_{i_2}-1}\prod_{t\neq i_1,i_2}(1-\lambda_t)+\cdots\Big\}p\left(z^d\right)\\
&= p\left(x^d\right)\prod_{t=1}^{r}(1-\lambda_t)+\sum_{z^d\in S^d}\sum_{s=1}^{r}\frac{\lambda_s}{c_s-1}\mathbf{1}_s(x^d,z^d)\prod_{t\neq s}^{r}(1-\lambda_t)p\left(z^d\right)\\
&\quad+\sum_{z^d\in S^d}\sum_{1\le i_1<i_2\le r}\frac{\lambda_{i_1}\lambda_{i_2}}{(c_{i_1}-1)(c_{i_2}-1)}\mathbf{1}_{i_1,i_2}(x^d,z^d)\prod_{t\neq i_1,i_2}(1-\lambda_t)p\left(z^d\right)+\cdots\\
&= p\left(x^d\right)+\sum_{s=1}^{r}\left(\sum_{z^d\in S^d}\frac{1}{c_s-1}\mathbf{1}_s(x^d,z^d)p\left(z^d\right)-p\left(x^d\right)\right)\lambda_s+O\left(\sum_{s=1}^{r}\lambda_s^2\right)\\
&= p\left(x^d\right)+\sum_{s=1}^{r}B_{p,s}\lambda_s+O\left(\sum_{s=1}^{r}\lambda_s^2\right).
\end{aligned}
$$

We have used the Geometric Inequality $\prod\limits_{i=1}^{n}a_i\le\sum\limits_{i=1}^{n}a_i^n/n$, for $a_i\ge 0, i=1,\cdots,n$ to obtain the form of the residual term in the previous deduction.

Now, we will get the form of $Var\left(\hat{p}(x)\right)$. Since the X_i's are i.i.d, $\hat{p}\left(x^d\right)=\frac{1}{n}\sum\limits_{i=1}^{n}L\left(X_i^d,x^d,\lambda\right)$, we have

$$
\begin{aligned}
Var\left(\hat{p}\left(x^d\right)\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}L\left(X_i^d,x^d,\lambda\right)\right)=\frac{1}{n}Var\left(L\left(X_1^d,x^d,\lambda\right)\right)\\
&= \frac{1}{n}\left[E\left(L^2\left(X_1^d,x^d,\lambda\right)\right)-\left(E\left(L\left(X_1^d,x^d,\lambda\right)\right)\right)^2\right]\\
&= \frac{1}{n}\Bigg\{p\left(x^d\right)\prod_{t=1}^{r}(1-\lambda_t)^2+\sum_{z^d\in S^d}\sum_{s=1}^{r}\frac{\lambda_s^2}{(c_s-1)^2}\mathbf{1}_s(x^d,z^d)\prod_{t\neq s}^{r}(1-\lambda_t)^2p\left(z^d\right)+\cdots\\
&\quad-\left[p\left(x^d\right)\prod_{t=1}^{r}(1-\lambda_t)+\sum_{z^d\in S^d}\sum_{s=1}^{r}\frac{\lambda_s}{c_s-1}\mathbf{1}_s(x^d,z^d)\prod_{t\neq s}^{r}(1-\lambda_t)p\left(z^d\right)+\cdots\right]^2\Bigg\}\\
&= \frac{p\left(x^d\right)\left(1-p\left(x^d\right)\right)}{n}+O\left(n^{-1}\sum_{s=1}^{r}\lambda_s\right).
\end{aligned}
$$

First, we show that $E\left[\hat{m}\left(x^d\right)\right] = O\left(\sum\limits_{s=1}^{r}\lambda_s\right)$.

$$\begin{aligned}
\hat{m}\left(x^d\right) &= \left[\hat{g}\left(x^d\right) - g\left(x^d\right)\right]\hat{p}\left(x^d\right) = n^{-1}\sum_{i=1}^{n} Y_i L\left(X_i^d, x^d, \lambda\right) - g\left(x^d\right)\hat{p}\left(x^d\right) \\
&= n^{-1}\sum_{i=1}^{n}\left(g\left(X_i^d\right) + u_i\right) L\left(X_i^d, x^d, \lambda\right) - g\left(x^d\right) n^{-1}\sum_{i=1}^{n} L\left(x_i^d, x^d, \lambda\right) \\
&= n^{-1}\sum_{i=1}^{n}\left(g\left(X_i^d\right) - g\left(x^d\right)\right) L\left(X_i^d, x^d, \lambda\right) + n^{-1}\sum_{i=1}^{n} u_i L\left(X_i, x^d, \lambda\right) \\
&= \hat{m}_1\left(x^d\right) + \hat{m}_2\left(x^d\right).
\end{aligned}$$

$$\begin{aligned}
E\left(\hat{m}_1\left(x^d\right)\right) &= E\left(n^{-1}\sum_{i=1}^{n}\left(g\left(X_i^d\right) - g\left(x^d\right)\right) L\left(X_i^d, x^d, \lambda\right)\right) \\
&= E\left(\left(g(X_1^d) - g\left(x^d\right)\right) L\left(X_1^d, x^d, \lambda\right)\right) \\
&= \sum_{z^d \in S^d}\left(g\left(z^d\right) - g\left(x^d\right)\right) L\left(z^d, x^d, \lambda\right) p\left(z^d\right) \\
&= \sum_{z^d \in S^d,\ z^d \neq x^d}\left(g\left(z^d\right) - g\left(x^d\right)\right) L\left(z^d, x^d, \lambda\right) p\left(z^d\right) \\
&= \sum_{z^d \in S^d,\ z^d \neq x^d}\sum_{s=1}^{r}\left(g\left(z^d\right) - g\left(x^d\right)\right)\lambda_s \mathbf{1}_s(x^d, z^d) p\left(z^d\right) + \cdots = O\left(\sum_{s=1}^{r}\lambda_s\right).
\end{aligned}$$

$$\begin{aligned}
E\left(\hat{m}_2\left(x^d\right)\right) &= E\left(n^{-1}\sum_{i=1}^{n} u_i L\left(X_i^d, x^d, \lambda\right)\right) = E\left(u_1 L\left(X_1^d, x^d, \lambda\right)\right) \\
&= E\left[E\left(u_1 L\left(X_1^d, x^d, \lambda\right)|X_1^d\right)\right] = E\left[L\left(X_1^d, x^d, \lambda\right) E\left(u_1|X_1^d\right)\right] = 0.
\end{aligned}$$

We use the Law of Iterated Expectations to get the third equality.

Hence, $E\left[\hat{m}\left(x^d\right)\right] = O\left(\sum\limits_{s=1}^{r}\lambda_s\right)$.

Second, we show that $Var\left[\hat{m}\left(x^d\right)\right] = O\left(n^{-1}\right)$. We know that X_i and u_i are uncorrelated. Hence, $Var\left[\hat{m}\left(x^d\right)\right] = Var\left[\hat{m}_1\left(x^d\right) + \hat{m}_2\left(x^d\right)\right] = Var\left[\hat{m}_1\left(x^d\right)\right] + Var\left[\hat{m}_2\left(x^d\right)\right]$.

$$
\begin{aligned}
Var\left[\hat{m}_1\left(x^d\right)\right] &= n^{-1}Var\left[\left(g(X_1^d) - g\left(x^d\right)\right)L\left(X_1^d, x^d, \lambda\right)\right] \\
&= n^{-1}\left\{E\left[\left(g(X_1^d) - g\left(x^d\right)\right)^2 L^2\left(X_1^d, x^d, \lambda\right)\right] - \left[E\left[\left(g(X_1^d) - g\left(x^d\right)\right)L\left(X_1^d, x^d, \lambda\right)\right]\right]^2\right\} \\
&= n^{-1}\Bigg\{\sum_{z^d \in S^d,\ z^d \neq x^d}\sum_{s=1}^{r}\left(g\left(z^d\right) - g\left(x^d\right)\right)^2 \lambda_s^2 \mathbf{1}_s(x^d, z^d) p\left(z^d\right) + \cdots \\
&\quad - \left(\sum_{z^d \in S^d,\ z^d \neq x^d}\sum_{s=1}^{r}\left(g\left(z^d\right) - g\left(x^d\right)\right)\lambda_s \mathbf{1}_s(x^d, z^d) p\left(z^d\right) + \cdots\right)^2\Bigg\} \\
&= O\left(n^{-1}\right).
\end{aligned}
$$

$$
\begin{aligned}
Var\left[\hat{m}_2\left(x^d\right)\right] &= n^{-1}Var\left[u_1 L\left(X_1^d, x^d, \lambda\right)\right] = n^{-1}E\left(u_1 L\left(X_1^d, x^d, \lambda\right)\right)^2 \\
&= n^{-1}E\left[E\left(u_1^2 L^2\left(X_1^d, x^d, \lambda\right)\middle| X_1^d\right)\right] = n^{-1}E\left(\sigma^2(X_1^d) L^2\left(X_1^d, x^d, \lambda\right)\right) \\
&= n^{-1}\sigma^2\left(x^d\right)p\left(x^d\right) + \sum_{z^d \in S^d,\ z^d \neq x^d}\sum_{s=1}^{r}\sigma^2\left(z^d\right)\lambda_s^2 \mathbf{1}_s(x^d, z^d) p\left(z^d\right) + \cdots = O\left(n^{-1}\right).
\end{aligned}
$$

We use $E\left(u_1 L\left(X_1^d, x^d, \lambda\right)\right) = E\left(L\left(X_1^d, x^d, \lambda\right)E(u_1|X_1)\right) = 0$ to get the second equality. Hence, $Var\left[\hat{m}\left(x^d\right)\right] = O\left(n^{-1}\right)$.

Therefore, we get $E\left(\hat{m}\left(x^d\right)^2\right) = Var\left[\hat{m}\left(x^d\right)\right] + \left(E\left[\hat{m}\left(x^d\right)\right]\right)^2 = O\left(\sum_{s=1}^{r}\lambda_s^2 + n^{-1}\right)$.

Exercise 4.4.

Exercise 4.5. We substitute $\sum_{x^d}\left[p\left(x^d\right)\right]^2 = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}L_{ij}^{(2)}$ into equation (4.8) (Li and Racine (2007, page 129)) to obtain

$$
\begin{aligned}
CV(\lambda) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}L_{ij}^{(2)} - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j \neq i}^{n}L_{ij} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}L_{ii}^{(2)} + \frac{1}{n(n-1)}\sum_{i=1}^{n}\left[L_{ij}^{(2)} - 2L_{ij}\right] - \frac{1}{n^2(n-1)}\sum_{j \neq i}^{n}L_{ij}^{(2)}. \qquad (4.1)
\end{aligned}
$$

We will first compute $E\left[L_{ii}^{(2)}\right]$, $E\left[L_{ij}^{(2)}\right]$ and $E\left[L_{ij}\right]$ $(j \neq i)$ Note that $L_{ij}^{(2)} = \sum_{x^d} L_{\lambda,ix}L_{\lambda,jx}$, we have

$$
\begin{aligned}
E\left[L_{ii}^{(2)}\right] &= \sum_{x^d}\sum_{z^d}p\left(z^d\right)L_{\lambda,zx}^2 \\
&= (1-\lambda)^2\sum_{x^d}p\left(x^d\right) + \frac{\lambda^2}{(c-1)^2}\sum_{x_d}\sum_{z^d \neq x^d}p\left(z^d\right) \\
&= (1-\lambda)^2 + O\left(\lambda^2\right) = 1 - 2\lambda + O\left(\lambda^2\right). \qquad (4.2)
\end{aligned}
$$

Define $\bar{p}\left(x^d\right) = \frac{1}{c-1}\sum_{z^d \neq x^d} p\left(z^d\right)$. We have

$$\begin{aligned}
E\left[L_{ij}^{(2)}\right] &= \sum_{x^d}\sum_{x_i^d}\sum_{x_j^d} p\left(x_i^d\right) p\left(x_j^d\right) L_{\lambda,ix}L_{\lambda,jx} \\
&= \sum_{x^d} p\left(x^d\right)^2 (1-\lambda)^2 + \sum_{x^d}\sum_{x_i^d \neq x^d} p\left(x_i^d\right)^2 \frac{\lambda^2}{(c-1)^2} \\
&\quad + 2\sum_{x^d}\sum_{x_j^d \neq x^d} p\left(x^d\right) p\left(x_j^d\right)\left(\frac{(1-\lambda)\lambda}{c-1}\right) + \sum_{x^d}\sum_{x_i^d \neq x^d}\sum_{x_j^d \neq x_i^d, x^d} p\left(x_i^d\right) p\left(x_j^d\right)\frac{\lambda^2}{(c-1)^2} \\
&= \left(1-2\lambda+\lambda^2\right)\sum_{x^d} p\left(x^d\right)^2 + \lambda^2\sum_{x^d}\left[\frac{1}{c-1}\sum_{x_i^d \neq x^d} p\left(x_i^d\right)\right]^2 \\
&\quad + 2(1-\lambda)\lambda\sum_{x^d} p\left(x^d\right)\left[\frac{1}{c-1}\sum_{x_i^d \neq x^d} p\left(x_i^d\right)\right] \\
&= \sum_{x^d} p\left(x^d\right)^2 + 2\lambda\left[\sum_{x^d} p\left(x^d\right)\bar{p}\left(x^d\right) - p\left(x^d\right)^2\right] + \lambda^2\sum_{x^d}\left[p\left(x^d\right)^2 - \bar{p}\left(x^d\right)\right]^2. \qquad (4.3)
\end{aligned}$$

$$\begin{aligned}
E\left[L_{ij}\right] &= \sum_{x_i^d}\sum_{x_j^d} p\left(x_i^d\right) p\left(x_j^d\right) L_{\lambda,x_ix_j} = \sum_{x^d} p\left(x^d\right)^2 (1-\lambda) + \sum_{x^d} p\left(x^d\right)\sum_{z^d \neq x^d} p\left(z^d\right)\frac{\lambda}{c-1} \\
&= \sum_{x^d} p\left(x^d\right)^2 + \lambda\sum_{x_d}\left[p\left(x^d\right)\bar{p}\left(x^d\right) - p\left(x^d\right)^d\right]. \qquad (4.4)
\end{aligned}$$

Combining (4.1) to (4.4) we obtain

$$E\left[CV(\lambda)\right] = D_1\lambda^2 - D_2\lambda n^{-1} + o\left(\lambda^2 + \lambda n^{-1}\right) + \text{terms unrelated to } \lambda,$$

where $D_1 = \sum_{x^d}\left[p\left(x^d\right) - \bar{p}\left(x^d\right)\right]^2$ and $D_2 = 2$, both are positive constants as claimed.

Exercise 4.6.

Exercise 4.7. Let $\hat{g}_i = \hat{g}_{-i}(X_i)$ and using $Y_i = g_i + u_i$, we have

$$CV_\lambda = n^{-1}\sum_i (Y_i - \hat{g}_i)^2 = n^{-1}\sum_i (g_i - \hat{g}_i)^2\hat{p}_i^2/\hat{p}_i^2 + 2n^{-1}\sum_i u_i(g_i - \hat{g}_i)\hat{p}_i/\hat{p}_i + n^{-1}\sum_i u_i^2. \qquad (4.5)$$

Since the third term is unrelated to λ, then minimizing CV_λ is equivalent to minimizing $CV_{\lambda,0}$:

$$CV_{\lambda,0} = n^{-1}\sum_i (g_i - \hat{g}_i)^2\hat{p}_i^2/\hat{p}_i^2 + 2n^{-1}\sum_i u_i(g_i - \hat{g}_i)\hat{p}_i/\hat{p}_i. \qquad (4.6)$$

Using $\hat{g}_i = n^{-1}\sum_{j\neq i} Y_j L_{ij}/\hat{p}_i$, $Y_j = g_j + u_j$, and $L_{ij} = L(x_i, x_j, \lambda)$, we have

$$
\begin{aligned}
CV_{\lambda,0} &= n^{-3}\sum_i\sum_{j\neq i}\sum_{l\neq i}(g_i - Y_j)(g_i - Y_l)L_{ij}L_{il}/\hat{p}_i^2 + 2n^{-2}\sum_i\sum_{j\neq i}u_i(g_i - Y_j)L_{ij}/\hat{p}_i \\
&= \left\{\frac{1}{n^3}\sum_i\sum_{j\neq i}\sum_{l\neq i}(g_i - g_j)(g_i - g_l)L_{ij}L_{il}/\hat{p}_i^2\right\} \\
&\quad + \left\{\frac{1}{n^3}\sum_i\sum_{j\neq i}\sum_{l\neq i}u_j u_l L_{ij}L_{il}/\hat{p}_i^2 - \frac{2}{n^2}\sum_i\sum_{j\neq i}u_i u_j L_{ij}/\hat{p}_i\right\} \\
&\quad + 2\left\{n^{-2}\sum_i\sum_{j\neq i}u_i(g_i - g_j)L_{ij}/\hat{p}_i - n^{-3}\sum_i\sum_{j\neq i}\sum_{l\neq i}(g_i - g_j)u_l L_{ij}L_{il}/\hat{p}_i^2\right\} \\
&\equiv S_1 + S_2 + 2S_3,
\end{aligned} \tag{4.7}
$$

In Lemmas 4.1 to 4.3 we prove that ($\mathcal{D}$ is the support of X)

$$
S_1 = \sum_{x\in\mathcal{D}}\left[\lambda\left(\sum_{z\in\mathcal{D}}p(z)\mathbf{1}_{x\neq z}(g(x) - g(z))\right)\right]^2 p(x)^{-1} + o_p\left(\lambda^2\right), \tag{4.8}
$$

where $\mathbf{1}_{x\neq z}$ is an indicator function which equals 1 if $x \neq z$, and zero otherwise.

$$
S_2 = -\frac{1}{n}\lambda\left[A + Z_{1n}\right] + o_p\left(\lambda^2 + n^{-1}\lambda\right) + \text{terms unrelated to } \lambda, \tag{4.9}
$$

where A is a positive constant and Z_{1n} is a zero mean $O_p(1)$ random variable.

$$
S_3 = \frac{1}{n}\lambda Z_{2n} + o_p\left(n^{-1}\lambda + \lambda^2\right) + \text{terms unrelated to } \lambda, \tag{4.10}
$$

where Z_{2n} is a zero mean $O_p(1)$ random variable. Therefore, (4.8), (4.9) and (4.10) lead to

$$
CV_{\lambda,0} = \sum_{x\in\mathcal{D}}\left[\lambda\sum_{z\in\mathcal{D}}\mathbf{1}_{x\neq z}p(z)\left(g(x) - g(z)\right)\right]^2 p(x)^{-1} - \frac{1}{n}\lambda\left(A + Z_n\right) + (s.o.), \tag{4.11}
$$

where $Z_n = Z_{1n} - 2Z_{2n}$ and $(s.o.) = o_p\left(n^{-1}\lambda + \lambda^2\right)$ + terms unrelated to λ.

Note that (4.8) can be written as $S_1 = \Omega\lambda^2 + o_p\left(\lambda^2\right)$, where

$$
\Omega = \sum_x\left[\sum_z \mathbf{1}_{x\neq z}p(z)(g(x) - g(z))\right]^2 p(x)^{-1} > 0
$$

provided that $g(.)$ is not a constant function. Thus, from (4.11) we have

$$
\frac{\partial CV_{\lambda,0}}{\partial\lambda} = 2\Omega\lambda - n^{-1}[A + Z_n] + (s.o.) \overset{set}{=} 0, \tag{4.12}
$$

(4.12) leads to $\hat{\lambda} = O_p\left(n^{-1}\right)$.

We will use the following identity to handle the random denominator,

$$\frac{1}{\hat{p}_i} = \frac{1}{p_i} + \frac{(p_i - \hat{p}_i)}{p_i^2} + \frac{(p_i - \hat{p}_i)^2}{p_i^2 \hat{p}_i}. \tag{4.13}$$

Defining $p_{i,0} = \frac{1}{n-1}\sum_{j\neq i} \mathbf{1}_{x_j = x_i}$ and $p_{i,1} = \frac{1}{n-1}\sum_{j\neq i} \mathbf{1}_{x_j \neq x_i}$, we have

$$\begin{aligned} p_i - \hat{p}_i &= p_i - \frac{1}{n-1}\sum_{j\neq i} L_{ij} = p_i - \frac{1}{n-1}\sum_{j\neq i}\left[\mathbf{1}_{x_j=x_i} + \lambda \mathbf{1}_{x_j\neq x_i}\right] \\ &= (p_i - p_{i,0}) - \lambda p_{i,1} = O_p\left(n^{-1/2}\right) + O_p(\lambda), \end{aligned} \tag{4.14}$$

the last equality following because $\max_{1\leq i\leq n}|p_i - p_{i,0}| \leq \sup_{x\in\mathcal{D}}\left|p(x) - n^{-1}\sum_i \mathbf{1}_{x_i=x}\right| + O\left(n^{-1}\right) = O_p\left(n^{-1/2}\right)$ (since $\mathcal{D}$ is a finite set) and $\max_{1\leq i\leq n} p_{i,1} = O_p(1)$.

Substituting (4.14) into (4.13), we get

$$\frac{1}{\hat{p}_i} = \frac{1}{p_i} + \frac{(p_i - p_{i,0})}{p_i^2} - \lambda\frac{p_{i,1}}{p_i^2} + O_p\left(n^{-1}\right) + O_p\left(n^{-1/2}\lambda + \lambda^2\right). \tag{4.15}$$

The above $O_p\left(n^{-1}\right)$ term comes from $\left(p_i - \tilde{p}_{i,0}\right)^2/p_i^3$ and is unrelated to λ. From (4.15) we get

$$\frac{1}{\hat{p}_i^2} = \frac{1}{p_i^2} + 2\frac{(p_i - p_{i,0})}{p_i^3} - 2\lambda\frac{p_{i,1}}{p_i^3} + O_p\left(n^{-1}\right) + O_p\left(n^{-1/2}\lambda + \lambda^2\right), \tag{4.16}$$

where again the $O_p\left(n^{-1}\right)$ term is unrelated to λ. Both (4.15) and (4.16) will be used to handle the random denominator in the proofs below.

Lemma 4.1. $S_1 = \sum_{x\in\mathcal{D}}\left[\lambda\left(\sum_{z\in\mathcal{D}}\mathbf{1}_{x\neq z}p(z)(g(x) - g(z))\right)\right]^2 p(x)^{-1} + o_p\left(\lambda^2\right)$.

Define S_1^0 the same way as S_1 except that $\hat{p}_i^{-2}$ is replaced by p_i^{-2}. That is,

$$S_1^0 \stackrel{def}{=} \frac{1}{n^3}\sum\sum_{i\neq j}(g_i - g_j)^2 L_{ij}^2/p_i^2 + \frac{1}{n^3}\sum\sum\sum_{i\neq j\neq l}(g_i - g_j)(g_i - g_l)L_{ij}L_{il}/p_i^2 = S_{1a} + S_{1b}.$$

It can be shown that S_{1a} is asymptotically negligible compared to S_{1b}. S_{1b} can be written as a third order U-statistic. Then by the U-statistic H-decomposition one can show that

$$S_{1b} = E\left[S_{1b}\right] + (s.o.). \tag{4.17}$$

By noting that $(g_i - g_j)\mathbf{1}_{x_j=x_i} = 0$, we obtain

$$E\left[(g_i - g_j)\, L_{ji}|x_j = x\right] = \lambda\sum_{z\in\mathcal{D}}\mathbf{1}_{x\neq z}p\left(z\right)\left(g\left(x\right) - g\left(z\right)\right) + O\left(\lambda^2\right). \tag{4.18}$$

Hence, we have

$$E\left[S_{1b}\right] = E\left[\left\{E\left[(g_i - g_j)L_{ji}|x_i\right]\right\}^2/p_i^2\right] = \sum_{x\in\mathcal{D}}\left[\lambda\sum_{z\in\mathcal{D}}\mathbf{1}_{x\neq z}p(z)\left(g(x) - g(z)\right)\right]^2 p(x)^{-1} + o\left(\lambda^2\right). \tag{4.19}$$

By (4.17) and (4.19) we have

$$S_1 = \sum_{x\in\mathcal{D}} \left[\sum_{s=1}^{r} \lambda_s \sum_{z\in\mathcal{D}} \mathbf{1}_{x\neq z} p(z)\left(g(x) - g(z)\right) \right]^2 p(x)^{-1} + o_p\left(\lambda^2\right). \tag{4.20}$$

Lemma 4.2. *$S_2 = -n^{-1}\lambda A + n^{-1}\lambda Z_{1n} + o_p\left(\lambda^2 + \lambda^{-1}\right) +$ terms unrelated to λ, where $A > 0$ is a positive constant, and Z_{1n} is a zero mean $O_p(1)$ random variable defined in the proof below.*

$$\begin{aligned} S_2 &= n^{-3} \sum\sum_{j\neq i} u_j^2 L_{ij}^2/\hat{p}_i^2 + n^{-3}\sum\sum\sum_{i\neq j\neq l} u_j u_l L_{ij} L_{il}/\hat{p}_i^2 - 2n^{-2}\sum\sum_{j\neq i} u_i u_j L_{ij}/\hat{p}_i \\ &\equiv S_{2a} + S_{2b} - 2S_{2c}. \end{aligned}$$

Using (4.16) and noting that $L_{ij,\lambda}^2 = O\left(\lambda^2\right)$ if $x_j \neq x_i$, we have

$$\begin{aligned} S_{2a} &= n^{-3}\sum\sum_{j\neq i} \mathbf{1}_{x_j=x_i} u_j^2/\hat{p}_i^2 + O_p\left(n^{-1}\lambda^2\right) \\ &= n^{-3}\sum\sum_{j\neq i} \mathbf{1}_{x_j=x_i} u_j^2 \left[1/p_i^2 + 2(p_i - p_{i,0})/p_i^3 - 2\lambda p_{i,1}/p_i^3\right] + O\left(n^{-3/2}\lambda + n^{-1/2}\lambda^2\right) \\ &= -2\frac{\lambda}{n}\left[n^{-2}\sum\sum_{j\neq i} \mathbf{1}_{x_j=x_i} u_j^2 p_{i,1}/p_i^3\right] + O_p\left(n^{-3/2}\lambda + n^{-1/2}\lambda^2\right) + \text{terms unrelated to } \lambda, \\ &\equiv -n^{-1}\lambda A + O\left(n^{-3/2}\lambda + n^{-1/2}\lambda^2\right) + \text{terms unrelated to } \lambda, \end{aligned}$$

where $A = 2E\left[\mathbf{1}_{x_j=x_i} u_j^2 p_{i,1}/p_i^2\right]$ is a positive constant, and we have used the fact that

$$2n^{-2}\sum\sum_{j\neq i} \mathbf{1}_{x_j=x_i} u_j^2 p_{i,1}/p_i^3 = A + O_p\left(n^{-1/2}\right). \tag{4.21}$$

(4.21) follows from the U-statistic H-decomposition because $2n^{-2}\sum\sum_{j\neq i}\mathbf{1}_{x_j=x_i} u_j^2 p_{i,1}/p_i^3$ can be written as a second order U-statistic.

$$\begin{aligned} S_{2b} &= n^{-3}\sum\sum\sum_{i\neq j\neq l} u_j u_l \left[\mathbf{1}(x_j = x_i) + \lambda \mathbf{1}_{x_j\neq x_i}\right]\left[\mathbf{1}_{x_l=x_i} + \lambda\mathbf{1}_{x_l\neq x_i}\right] \\ &\qquad \times \left[1/p_i^2 + 2(p_i - p_{i,0})/p_i^3 - 2\lambda p_{i,1}/p_i^3\right] + O_p\left(n^{-1}\lambda^2\right) \\ &= n^{-1}\lambda 2n^{-2}\sum\sum\sum_{i\neq j\neq l} u_j u_l \Big[\mathbf{1}_{x_j=x_i}\mathbf{1}_{x_l\neq x_i} + \mathbf{1}(x_l = x_i)\mathbf{1}_{x_j\neq x_i} \\ &\qquad - 2\times \mathbf{1}_{x_j=x_i}\mathbf{1}_{x_l=x_i} p_{i,1} p_i^{-1}\Big]/p_i^2 + O_p\left(n^{-3/2}\lambda + n^{-1/2}\lambda^2\right) + \text{ terms unrelated to } \lambda \\ &= n^{-1}\lambda Z_{3n} + O_p\left(n^{-3/2}\lambda + n^{-1/2}\lambda^2\right) + \text{ terms unrelated to } \lambda, \end{aligned}$$

where $Z_{3n} = \frac{2}{n^2}\sum\sum\sum_{i\neq j\neq l}\frac{u_j u_l}{p_i^2}\left[\mathbf{1}_{x_j=x_i}\mathbf{1}_{x_l\neq x_i} + \mathbf{1}_{x_l=x_i}\mathbf{1}(x_j\neq x_i) - 2\times\mathbf{1}_{x_j=x_i}\mathbf{1}_{x_l=x_i}\times\frac{p_{i,1}}{p_i}\right]$ is a zero mean $O_p(1)$ random variable, and we have also used the fact that $p_i - p_{i,0} = O_p\left(n^{-1/2}\right)$.

Letting $\zeta_n = n^{-1}\lambda^2 + n^{-3/2}\lambda +$ terms unrelated to λ, we have

$$
\begin{aligned}
S_{2c} &= n^{-2}\sum\sum_{j\neq i} u_i u_j\left[\mathbf{1}_{x_j=x_i} + \lambda\mathbf{1}_{x_j\neq x_i}\right]\left[1/p_i + (p_i - p_{i,0})/p_i^2 - \lambda p_{i,1}/p_i^2\right] + O_p(\zeta_n) \\
&= n^{-1}\sum_{s=1}^{r}\lambda n^{-1}\sum\sum_{j\neq i} u_i u_j\left[\mathbf{1}_{x_j\neq x_i}/p_i - \mathbf{1}_{x_j=x_i}p_{i,1}/p_i^2\right] + O_p(\zeta_n) \\
&= n^{-1}\lambda Z_{4n} + O_p\left(n^{-1}\lambda^2 + n^{-3/2}\lambda\right) + \text{ terms unrelated to } \lambda,
\end{aligned}
$$

where $Z_{4n} = n^{-1}\sum\sum_{j\neq i} u_i u_j[\mathbf{1}(x_j\neq x_i)/p_i - \mathbf{1}_{x_j=x_i}p_{i,1}/p_i^2]$ is a zero mean $O_p(1)$ random variable. The term associated with $p_i - p_{i,0}$ is of order $O_p(n^{-3/2}\lambda)$ because $\max_{1\leq i\leq n}|p_i - p_{i,0}| = O_p\left(n^{-1/2}\right)$.

Summarizing the above we have shown that

$$
S_2 = S_{2a} + S_{2b} - 2S_{2c} = -n^{-1}\lambda A + n^{-1}\lambda Z_{1n} + +O_p(\zeta_n) \tag{4.22}
$$

where $Z_{1n} = Z_{3n} - 2Z_{4n}$ is a zero mean $O_p(1)$ random variable.

Lemma 4.3.

$$
S_3 = n^{-1}\lambda Z_{2n} + o_p\left(\lambda^2 + n^{-1}\lambda\right),
$$

where Z_{2n} is a zero mean $O_p(1)$ random variable defined in the proof below.

$$
\begin{aligned}
S_3 = n^{-2}\sum_i\sum_{j\neq i} u_i(g_i - g_j)L_{ij}/\hat{p}_i - n^{-3}\sum\sum\sum_{i\neq j\neq l}(g_i - g_j)u_l L_{ij}L_{il}/\hat{p}_i^2 \\
- n^{-3}\sum_i\sum_{j\neq i}(g_i - g_j)u_j L_{ij}^2/\hat{p}_i^2 \equiv S_{3a} - S_{3b} - S_{3c}.
\end{aligned}
$$

Letting $\xi_n = \lambda^2 n^{-1/2} + \lambda n^{-3/2}+$ terms unrelated to λ, then using (4.15) and noting that $(g_i - g_j)\mathbf{1}_{x_j=x_i} = 0$, we have

$$
\begin{aligned}
S_{3a} &= n^{-2}\sum\sum_{j\neq i} u_i(g_i - g_j)\left[0 + \sum_{s=1}^{r}\lambda\mathbf{1}_{x_j\neq x_i}\right]\left[1/p_i + (p_i - p_{i,0})/p_i^2\right] + O_p(\xi_n) \\
&= \lambda n^{-2}\sum\sum_{j\neq i}\mathbf{1}_{x_j\neq x_i}u_i(g_i - g_j)/p_i + \lambda n^{-2}\sum\sum_{j\neq i}\mathbf{1}_{x_j\neq x_i}u_i(g_i - g_j)(p_i - p_{i,0})/p_i^2 + O_p(\xi_n) \\
&= S_{3a,1} + S_{3a,2} + O_p(\xi_n).
\end{aligned} \tag{4.23}
$$

Next, we consider S_{3b}. Again noting that $(g_i - g_j)\mathbf{1}_{x_j=x_i} = 0$ and using (4.16), we have

$$\begin{aligned}
S_{3b} &= n^{-3}\sum\sum\sum_{i\neq j\neq l} u_l(g_i-g_j)\left[0+\lambda\mathbf{1}(x_j\neq x_i)\right]\left[\mathbf{1}_{x_l=x_i}\right]\left[1/p_i^2+2(p_i-p_{i,0})/p_i^3\right]+O_p\left(\xi_n\right)\\
&= \lambda n^{-3}\sum\sum\sum_{l\neq j\neq i}\mathbf{1}_{x_j\neq x_i}\mathbf{1}_{x_l=x_i}u_l(g_i-g_j)/p_i^2\\
&\quad + 2\lambda n^{-3}\sum\sum\sum_{l\neq j\neq i}\mathbf{1}(x_j\neq x_i)\mathbf{1}_{x_l=x_i}u_l(g_i-g_j)(p_i-p_{i,0})/p_i^3+O_p\left(\xi_n\right)\\
&\equiv S_{3b,1}+2S_{3b,2}+O_p\left(n^{-1/2}\lambda^2+n^{-3/2}\lambda\right). \qquad (4.24)
\end{aligned}$$

Note that $\max_{1\leq i\leq n}|p_i - p_{i,0}| = O_p\left(n^{-1/2}\right)$. It is easy to see that both $S_{3a,2}$ and $S_{3b,2}$ are of order $O_p\left(\lambda n^{-1}\right)$. Although $S_{3a,1}$ and $S_{3b,1}$ are both of order $O_p\left(\lambda n^{-1/2}\right)$, we will show that $S_{3a,1} - S_{3b,1} = O_p\left(\lambda n^{-1}\right)$. To show this, we need to re-write $S_{3b,1}$ in a form similar to $S_{3a,1}$,

$$\begin{aligned}
S_{3b,1} &= \lambda n^{-3}\sum_i\sum_{j\neq i}\sum_{l\neq i,l\neq j}\mathbf{1}_{x_j\neq x_i}\mathbf{1}_{x_l=x_i}u_l(g_i-g_j)/p_i^2\\
&= \lambda n^{-3}\sum_i\sum_{j\neq i}\sum_{l\neq i,l\neq j}\mathbf{1}_{x_j\neq x_l}\mathbf{1}_{x_l=x_i}u_l(g_l-g_j)/p_l^2 \qquad (\text{since } x_l=x_i)\\
&= \lambda n^{-2}\sum_j\sum_{l\neq j}\left[\mathbf{1}_{x_j\neq x_l}u_l(g_l-g_j)/p_l^2\right]\left[n^{-1}\sum_{i\neq j,i\neq l}\mathbf{1}_{x_l=x_i}\right]\\
&= \lambda n^{-2}\sum_j\sum_{l\neq j}\mathbf{1}(x_j\neq x_l)u_l(g_l-g_j)p_{l,0}^*/p_l^2 \qquad \left(p_{l,0}^*\overset{def}{=}n^{-1}\textstyle\sum_{i=1,i\neq j,i\neq l}^n\mathbf{1}_{x_l=x_i}\right)\\
&= \lambda n^{-2}\sum_j\sum_{i\neq j}\mathbf{1}_{x_j\neq x_i}u_i(g_i-g_j)p_{i,0}/p_i^2+O_p\left(\lambda n^{-3/2}\right), \qquad (4.25)
\end{aligned}$$

where in the second equality we used $g_l = g_i$, $p_l = p_i$ because $x_l = x_i$ due to the restriction $\mathbf{1}(x_l = x_i)$. The third equality simply reorders the summations. The fourth equality follows from the definition of $p_{l,0}^*$, while in the last equality we used $\max_{1\leq l\leq n}|p_{l,0} - p_{l,0}^*| = O_p\left(n^{-1}\right)$ $\left(p_{l,0} = n^{-1}\sum_{i=1,i\neq l}\mathbf{1}_{x_l=x_i}\right)$ and we changed summation indices from (j,l) to (j,i).

Note that both $S_{3a,1}$ and $S_{3b,1}$ are of order $O_p\left(\lambda n^{-1/2}\right)$, but we have (using (4.25))

$$\begin{aligned}
S_{3a,1}-S_{3b,1} &= n^{-1}\lambda\left\{n^{-1}\sum\sum_{j\neq i}\mathbf{1}_{x_j\neq x_i}u_i(g_i-g_j)(p_i-p_{i,0})/p_i^2\right\}+O_p\left(n^{-3/2}\lambda\right)\\
&\equiv S_{3a,2}+O_p\left(n^{-3/2}\lambda\right), \qquad (4.26)
\end{aligned}$$

which is of the order of $O_p\left(\lambda n^{-1}\right)$ because $E\left[S_{3a,2}^2\right] = O\left(\lambda^2 n^{-2}\right)$ (since $\max_{1\leq i\leq n}|p_{i,0} - p_i| = O_p\left(n^{-1/2}\right)$). Finally, noting that $(g_i - g_j)\mathbf{1}_{x_j=x_i} = 0$, and that $\mathbf{1}(x_j \neq x_i)L_{ij}^2 = O\left(\lambda^2\right)$, we have

$$S_{3c} = n^{-3}\sum\sum_{j\neq i}\mathbf{1}_{x_j\neq x_i}(g_i-g_j)u_jL_{ij}^2/\hat{p}_i^2 = O\left(n^{-1}\lambda^2\right). \qquad (4.27)$$

Now, combining (4.23), (4.24), (4.26), and (4.27), we obtain

$$\begin{aligned}
S_3 &= (S_{3a,1} - S_{3b,1}) + S_{3a,2} - 2S_{3b,2} + O_p(\xi_n) = 2S_{3a,2} - 2S_{3b,2} + O_p(\xi_n) \\
&= n^{-1}\lambda \left\{ \frac{2}{n} \sum \sum_{j \neq i} \mathbf{1}_{x_j \neq x_i} (g_i - g_j)(p_i - p_{i,0})/p_i^2 \left[u_i - \frac{1}{n} \sum_{l \neq i, l \neq j} \mathbf{1}_{x_l = x_i} u_l / p_i \right] \right\} + O_p(\xi_n) \\
&\equiv n^{-1}\lambda Z_{2n} + O_p\left(\lambda^2 n^{-1/2} + \lambda n^{-3/2}\right), \qquad (4.28)
\end{aligned}$$

where $Z_{2n} = \frac{2}{n} \sum \sum_{j \neq i} \mathbf{1}_{x_j \neq x_i} (g_i - g_j)(p_i - p_{i,0})/p_i^2 \left[u_i - \frac{1}{n} \sum_{l \neq i, l \neq j} \mathbf{1}_{x_l = x_i} u_l / p_i \right]$. Using

$$\max_{1 \leq i \leq n} |p_{i,0} - p_i| = O\left(n^{-1/2}\right),$$

it is easy to see that Z_{2n} is a zero mean $O_p(1)$ random variable.

Exercise 4.8.

Exercise 4.9.

(i) $\hat{m}(x) = \hat{m}_1(x) + \hat{m}_2(x)$, where $\hat{m}_1(x) = (nh_1 \dots h_q)^{-1} \sum_i (g(X_i) - g(x)) K_{ix}$ and $\hat{m}_2(x) = (nh_1 \dots h_q)^{-1} \sum_i u_i K_{ix}$. Obviously, $E(\hat{m}_2(x)) = 0$. Hence, $E(\hat{m}(x)) = E(\hat{m}_1(x))$, and

$$\begin{aligned}
E(\hat{m}_1(x)) &= (h_1 \dots h_q)^{-1} \sum_{x_1^d \in S^d} \int [g(x_1) - g(x)] W\left(\frac{x_1^c - x^c}{h}\right) L\left(x_1^d, x^d, \lambda\right) dx_1^c \\
&= (h_1 \dots h_q)^{-1} \int \left[g(x_1^c, x^d) - g(x)\right] W\left(\frac{x_1^c - x^c}{h}\right) dx_1^c \\
&\quad + (h_1 \dots h_q)^{-1} \sum_{x_1^d \in S^d} \mathbf{1}_s(x_1^d, x^d) \lambda_s \int [g(x_1) - g(x)] W\left(\frac{x_1^c - x^c}{h}\right) dx_1^c + O\left(|\lambda|^2\right) \\
&= (\kappa_2/2) \sum_{s=1}^{q} h_s^2 \left[g_{ss}(x) + 2g_s(x) f_s(x)/f(x)\right] \\
&\quad + \sum_{x_1^d \in S^d} \mathbf{1}_s(x_1^d, x^d) \lambda_s \left[g(x^c, x_1^d) - g(x)\right] + o\left(|h|^2 + |\lambda|\right) \\
&= \sum_{s=1}^{q} B_{1s}(x) f(x) h_s^2 + \sum_{s=1}^{r} B_{2s}(x) f(x) \lambda_s + o(|h|^2 + |\lambda|),
\end{aligned}$$

where the second term in the third equality follows since

$$\begin{aligned}
(h_1 \dots h_q)^{-1} \int [g(x_1) - g(x)] W\left(\frac{x_1^c - x^c}{h}\right) dx_1^c &= (h_1 \dots h_q)^{-1} \\
&\quad \times \int \left[g(x^c + hv, x_1^d) - g(x^c, x^d)\right] W(v) (h_1 \dots h_q) dv \\
&= \left[g(x^c, x_1^d) - g(x^c, x^d)\right] \left[\int W(v) dv\right] + O\left(|h|^2\right) \\
&= \left[g(x^c, x_1^d) - g(x^c, x^d)\right] + O\left(|h|^2\right).
\end{aligned}$$

Note that $g(x^c, x_1^d) - g(x^c, x^d) \neq 0$ since $x_1^d \neq x^d$ (because $||x_1^d - x^d|| = 1$ due to the factor $\mathbf{1}_s(x_1^d, x^d)$).

(ii) It can be shown that the leading term of $Var\left(\hat{m}(x)\right)$ comes from $Var\left(\hat{m}_2(x)\right)$. Also, using $L\left(X_i^d, x^d, \lambda\right) = \mathbf{1}(X_i^d = x^d) + O(|\lambda|)$, we have

$$\begin{aligned} Var\left(\hat{m}_2(x)\right) &= \left(nh_1^2 \dots h_q^2\right)^{-1} \sum_{x_1^d \in S^d} \int f(x_1)\sigma^2(x_1) W^2\left(\frac{x_1^c - x^c}{h}\right) L^2\left(x_1^d, x^d, \lambda\right) dx_1^c \\ &= \left(nh_1^2 \dots h_q^2\right)^{-1} \left\{\int f(x_1^c, x^d)\sigma^2(x_1^c, x^d) W^2\left(\frac{x_1^c - x^c}{h}\right) dx_1^c + O(|\lambda|)\right\} \\ &= \left(nh_1^2 \dots h_q^2\right)^{-1} \left\{\int f(x^c + hv, x^d)\sigma^2(x^c + hv, x^d) W^2(v)(h_1 \dots h_q) dv + O(|\lambda|)\right\} \\ &= (nh_1 \dots h_q)^{-1} \left\{\sigma^2(x^c, x^d) \int W^2(v) dv + O\left(|h|^2\right) O(|\lambda|)\right\} \\ &= \frac{\kappa^q \sigma^2(x) f(x)}{nh_1 \dots h_q}[1 + O\left(|h|^2 + |\lambda|\right)]. \end{aligned}$$

Similarly, one can show that $Var\left(\hat{m}_2(x)\right) = O\left(|h|^2 (nh_1 \dots h_q)^{-1}\right) = o\left((nh_1 \dots h_q)^{-1}\right)$. This completes the proof for (ii).

(iii) It is easy to show that $E(\hat{f}(x)) = f(x) + O(|h|^2 + |\lambda|)$ and $Var\left(\hat{f}(x)\right) = O\left((nh_1 \dots h_q)^{-1}\right)$. These results imply that $MSE(\hat{f}(x)) = O\left(\left(|h|^2 + |\lambda|\right)^2 + (nh_1 \dots h_q)^{-1}\right)$ and $\hat{f}(x) - f(x) = O_p\left(|h|^2 + |\lambda| + (nh_1 \dots h_q)^{-1/2}\right) = o_p(1)$.

Exercise 4.10.

Exercise 4.11. The hint in Li and Racine (2007, page 153) provides a solution to this problem.

Chapter 5

Conditional Density Estimation: Solutions

Exercise 5.1. There are some typos in Li and Racine (2007) regarding the statement of this exercise. In particular, $B_{20}(\bar{x}, y)$ was missing and $\hat{h}_s^2$ ($\hat{\lambda}_s$) should not appear in the definition of B_{1s} (B_{2s}). The correct expression should be:

$$\left(n\hat{h}_1 \dots \hat{h}_{q_1}\right)^{1/2} \left(\hat{g}(y|x) - g(y|x) - \sum_{s=1}^{q_1} B_{1s}(\bar{x}, y)\hat{h}_s^2 - \sum_{s=0}^{r_1} B_{2s}(\bar{x}, y)\hat{\lambda}_s\right) \xrightarrow{d} N\left(0, \sigma_g^2(\bar{x}, y)\right),$$

where

$$\begin{aligned}
B_{1s}(\bar{x}, y) &= \frac{1}{2}\kappa_2 \left[\frac{\bar{f}_{ss}(\bar{x}, y)}{\bar{\mu}(\bar{x})} - \frac{\bar{\mu}_{ss}(\bar{x})}{\bar{\mu}(\bar{x})}\bar{g}(y|\bar{x})\right], \\
B_{20}(\bar{x}, y) &= -\bar{g}(y|\bar{x}) + \frac{1}{c_0 - 1}\sum_{z^d \in \mathcal{S}_0^d} \mathbf{1}(z^d \neq y)\bar{g}(z^d|\bar{x}), \\
B_{2s}(\bar{x}, y) &= \frac{1}{c_s - 1}\sum_{\bar{v}^d \in \mathcal{S}_{\bar{x}}} \mathbf{1}_s(\bar{v}^d, \bar{x}_s^d) \left[\frac{\bar{f}(y, \bar{x}^c, \bar{v}^d)}{\bar{\mu}(\bar{x})} - \frac{\bar{\mu}(\bar{x}^c, \bar{v}^d)}{\bar{\mu}(\bar{x})}\bar{g}(y|\bar{x})\right], \\
\sigma_g^2(\bar{x}, y) &= \kappa^{q_1}\bar{g}(y|\bar{x})/\bar{\mu}(\bar{x}).
\end{aligned}$$

Proof. Under conditions given in Theorem 5.2 with a discrete random variable Y, we obtain a result similar to Theorem 5.2, namely that all irrelevant variables can be smoothed out asymptotically. Therefore, we will only consider the case where all variables are relevant in order to simplify the proof.

Let

$$\hat{f}(y, \bar{x}) = \frac{1}{n}\sum_{i=1}^n K_\gamma(\bar{x}, \bar{X}_i) l(y, Y_i, \lambda_0) \quad \text{and} \quad \hat{\mu}(\bar{x}) = \frac{1}{n}\sum_{i=1}^n K_\gamma(\bar{x}, \bar{X}_i),$$

where

$$K_\gamma(\bar{x}, \bar{X}_i) = W_h(\bar{x}^c, \bar{X}_i^c) L(\bar{x}^d, \bar{X}_i^d, \lambda),$$
$$W_h(\bar{x}^c, \bar{X}_i^c) = \prod_{s=1}^{q_1} \frac{1}{h_s} w\left(\frac{\bar{x}_s^c - \bar{X}_{is}^c}{h_s}\right),$$
$$L(\bar{x}^d, \bar{X}_i^d, \lambda) = \prod_{s=1}^{r_1} \left(\frac{\lambda_s}{c_s - 1}\right)^{\mathbf{1}(\bar{X}_{is}^d \neq \bar{x}_s^d)} (1 - \lambda_s)^{\mathbf{1}(\bar{X}_{is}^d = \bar{x}_s^d)},$$
$$l(y, Y_i, \lambda_0) = \left(\frac{\lambda_0}{c_0 - 1}\right)^{\mathbf{1}(Y_i \neq y)} (1 - \lambda_0)^{\mathbf{1}(Y_i = y)}.$$

We write

$$\hat{g}(y|\bar{x}) - \bar{g}(y|\bar{x}) = \frac{[\hat{g}(y|\bar{x}) - \bar{g}(y|\bar{x})]\,\hat{\mu}(\bar{x})}{\hat{\mu}(\bar{x})} \equiv \frac{\hat{m}(y, \bar{x})}{\hat{\mu}(\bar{x})}.$$

Then $E[\hat{m}(y, \bar{x})] = E[\hat{f}(y, \bar{x})] - \bar{g}(y|\bar{x})E[\hat{\mu}(x)]$ (because $\hat{g}(y|\bar{x})\hat{\mu}(\bar{x}) = \hat{f}(y, \bar{x})$). Following the approach used in the previous chapters (i.e., basic calculus and change of variables) we have that

$$E\left[\hat{f}(y, \bar{x})\right] = f(y, \bar{x}) - \lambda_0 f(y, \bar{x}) - \left(\sum_{s=1}^{r_1} \lambda_s\right) f(y, \bar{x}) + \frac{\lambda_0}{c_0 - 1} \sum_{z^d \in \mathcal{S}_0^d} \mathbf{1}(z^d \neq y) f(z^d, \bar{x})$$
$$+ \sum_{s=1}^{r_1} \frac{\lambda_s}{c_s - 1} \sum_{\bar{v}^d \in \mathcal{S}_{\bar{x}}} \mathbf{1}_s(\bar{v}^d, \bar{x}_s^d) f(y, \bar{x}^c, \bar{v}^d) + \frac{1}{2}\kappa_2 \sum_{s=1}^{q_1} h_s^2 f_{ss}(y, \bar{x}) + o\left(\sum_{s=1}^{q_1} h_s^2 + \sum_{s=0}^{r_1} \lambda_s\right).$$

The above result is quite easy to understand. When $(Y_i, \bar{X}_i^d) = (y, \bar{x}^d)$, the corresponding discrete kernel is $\prod_{s=0}^{r_1}(1 - \lambda_s) = 1 - \lambda_0 - \sum_{s=1}^{r_1} \lambda_s + O(|\lambda|^2)$, where $|\lambda|^2 = \sum_{s=0}^{r_1} \lambda_s^2$. When $(Y_i, \bar{X}_i^d) \neq (y, \bar{x}^d)$, the leading term is that for which only one component differs, say only the s^{th} components differ from each other. This gives the term associated with $\frac{\lambda_s}{c_s - 1}$ $(s = 0, \dots, r_1)$. The bias that arises due to the continuous $\bar{x}_s^c$ has the familiar order, h_s^2.

Next,

$$E\left[\hat{\mu}(\bar{x})\right] = \mu(\bar{x}) - \left(\sum_{s=1}^{r_1} \lambda_s\right) \mu(\bar{x}) + \sum_{s=1}^{r_1} \frac{\lambda_s}{c_s - 1} \sum_{\bar{v}^d \in \mathcal{S}_{\bar{x}}} \mathbf{1}_s(\bar{v}^d, \bar{x}^d) \mu(\bar{x}^c, \bar{v}^d)$$
$$+ \frac{1}{2}\kappa_2 \sum_{s=1}^{q_1} h_s^2 \mu_{ss}(\bar{x}) + o\left(\sum_{s=1}^{q_1} h_s^2 + \sum_{s=0}^{r_1} \lambda_s\right).$$

Hence,

$$E\left[\hat{m}(y, \bar{x})\right] = E[\hat{f}(y, \bar{x})] - \bar{g}(y|\bar{x})E[\hat{\mu}(\bar{x})]$$
$$= -\lambda_0 f(y, \bar{x}) + \frac{\lambda_0}{c_0 - 1} \sum_{z^d \in \mathcal{S}_0^d} \mathbf{1}(z^d \neq y) f(z^d, \bar{x})$$
$$\sum_{s=1}^{r_1} \frac{\lambda_s}{c_s - 1} \sum_{\bar{v}^d \in \mathcal{S}_{\bar{x}}} \mathbf{1}_s\left(\bar{v}^d, \bar{x}_s^d\right) \left[f(y, \bar{x}^c, \bar{v}^d) - \mu(\bar{x}^c, \bar{v}^d)\bar{g}(y|\bar{x})\right]$$
$$\frac{1}{2}\kappa_2 \sum_{s=1}^{q_1} h_s^2 \left[f_{ss}(y, \bar{x}) - \mu_{ss}(\bar{x})\bar{g}(y|\bar{x})\right] + o\left(\sum_{s=1}^{q_1} h_s^2 + \sum_{s=0}^{r_1} \lambda_s\right).$$

Also, it is easy to show that

$$Var\left[\hat{m}(y,\bar{x})\right] = Var\left[\hat{f}(y,\bar{x}) - \bar{g}(y|\bar{x})\hat{\mu}(x)\right]$$
$$= \frac{1}{nh_1\cdots h_{q_1}}\kappa^{q_1} f(y,\bar{x}) + o\left(\frac{1}{nh_1\cdots h_{q_1}} + \sum_{s=1}^{q_1} h_s^2 + \sum_{s=0}^{r_1}\lambda_s\right).$$

Note that

$$\hat{g}(y|\bar{x}) = \frac{\hat{m}(y,\bar{x})}{\hat{\mu}(\bar{x})} = \frac{\hat{m}(y,\bar{x})}{\mu(\bar{x})} + (s.o.) = \hat{g}_0(y|x) + (s.o.),$$

where $\hat{g}_0(y|\bar{x}) = \hat{m}(y,\bar{x})/\mu(\bar{x})$. The asymptotic distribution of $\hat{g}(y|\bar{x})$ is the same as that of $\hat{g}_0(y|\bar{x})$.

Obviously, $E[\hat{g}_0(y|x)] = E[\hat{m}(y,\bar{x})]/\mu(x)$ and $Var[\hat{g}_0(y|x)] = Var[\hat{m}(y,\bar{x})]/\mu(\bar{x})^2$. Therefore, by Liapunov's Central Limit Theorem, Slutsky's Lemma and noting that $g(y|x) = \bar{g}(y|\bar{x})$, we have

$$\left(n\hat{h}_1\dots\hat{h}_{q_1}\right)^{1/2}\left[\hat{g}(y|x) - g(y|x) - \sum_{s=1}^{q_1} B_{1s}(\bar{x},y)\hat{h}_s^2 - \sum_{s=0}^{r_1} B_{2s}(\bar{x},y)\hat{\lambda}_s\right] \xrightarrow{d} N\left(0,\sigma_g^2(\bar{x},y)\right),$$

where B_{1s} $(s=1,\dots,q_1)$, B_{20}, B_{2s} $(s=1,\dots,r_1)$ and $\sigma_g^2(\bar{x},y)$ are defined in the beginning of this solution. This completes the proof.

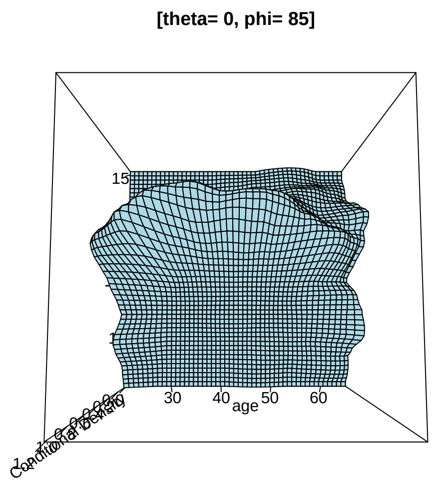
□

Exercise 5.2. Using the data underlying the local constant kernel estimate of an earnings profile (log income versus age) presented in Pagan and Ullah (1999, page 155) that we used in Exercise 2.11 (which is also part of the `np` package (Hayfield and Racine (2008))), generate the PDF of earnings conditional on age using least squares cross-validation.

```
R> data(cps71)
R> attach(cps71)
R> fhat <- npcdens(logwage~age,bwmethod="cv.ls")
R> summary(fhat)

Conditional Density Data: 205 training points, in 2 variable(s)
(1 dependent variable(s), and 1 explanatory variable(s))

                        logwage
Dep. Var. Bandwidth(s):   0.162
                         age
Exp. Var. Bandwidth(s): 2.71

Bandwidth Type: Fixed
Log Likelihood: -110

Continuous Kernel Type: Second-Order Epanechnikov
No. Continuous Explanatory Vars.: 1
No. Continuous Dependent Vars.: 1
```

Next, plot the resulting conditional PDF and compare your estimate with the conditional mean function you generated in Exercise 2.11. Can you readily visualize the conditional mean function from the conditional PDF function?

```
R> plot(fhat,view="fixed",phi=85)
```



Looking down on the estimated conditional density, you can certainly visualize the conditional mean running along the 'ridge' (i.e., the average value of logwages conditional on age). Note that logwage is on the 'vertical' axis, age on the 'horizontal'.

1

Chapter 6

Conditional CDF and Quantile Estimation: Solutions

Exercise 6.1. We will use the notation $f_{y|x}(y) = f(y|x)$ and $F_{y|x}(y) = F(y|x)$ to denote conditional PDF and CDF of $Y = y$ given $X = x$. With this notation we have (using $-\int_{\infty}^{-\infty} = \int_{-\infty}^{\infty}$)

$$
\begin{aligned}
E\left[G\left(\frac{y-Y_i}{h_0}\right)|X_i\right] &= \int_{-\infty}^{+\infty} G\left(\frac{y-y_i}{h_0}\right) f_{y|X_i}(y_i)dy_i \\
&= h_0 \int_{-\infty}^{+\infty} G(v) f_{y|X_i}(y-h_0v)dv \\
&= -\int_{-\infty}^{+\infty} G(v)dF_{y|X_i}(y-h_0v) \\
&= -G(v)F_{y|X_i}(y-h_0v)\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} w(v)F_{y|X_i}(y-h_0v)dv \\
&= 0 + \int_{-\infty}^{+\infty} w(v)\Big(F(y|X_i) - F_0(y|X_i)h_0v + \frac{1}{2}F_{00}(y|X_i)h_0^2v^2 + o(h_0^2)\Big)dv \\
&= F(y|X_i) + \frac{1}{2}\kappa_2 h_0^2 F_{00}(y|X_i) + o(h_0^2).
\end{aligned}
$$

Using integration by parts we have

$$
\begin{aligned}
E[G^2\left(\frac{y-Y_i}{h_0}\right)|X_i] &= \int_{-\infty}^{+\infty} G^2\left(\frac{y-y_i}{h_0}\right) f_{y|X_i}(y_i)dy_i \\
&= -h_0 \int_{+\infty}^{-\infty} G^2(v) f_{y|X_i}(y-h_0v)dv \\
&= h_0 \int_{-\infty}^{+\infty} G^2(v) f_{y|X_i}(y-h_0v)dv \quad (\text{since } -\textstyle\int_{+\infty}^{-\infty} = \int_{-\infty}^{\infty}) \\
&= -\int_{-\infty}^{+\infty} G^2(v) dF_{y|X_i}(y-h_0v) \\
&= -G^2(v)F_{y|X_i}(y-h_0v)\Big|_{-\infty}^{+\infty} + 2\int_{-\infty}^{+\infty} G(v)w(v)F_{y|X_i}(y-h_0v)dv \\
&= 0 + 2\int_{-\infty}^{+\infty} G(v)w(v)\Big(F_{y|X_i}(y) - F_{y|X_i,0}(y)h_0v + O(h_0^2)\Big)dv \\
&= 2\int_{-\infty}^{+\infty} G(v)w(v)\Big(F(y|X_i) - F_0(y|X_i)h_0v + O(h_0^2)\Big)dv \qquad (\text{since } F_{y|X_i}(y) = F(y|X_i)) \\
&= F(y|X_i) - h_0C_kF_0(y|X_i) + O(h_0^2),
\end{aligned}
$$

where $F_0(y|X_i) = \partial F(y|X_i)/\partial y$, $C_k = 2\int G(v)w(v)vdv$, and we have used $2\int G(v)w(v)dv = \int dG^2(v) = G^2(\infty) - G^2(-\infty) = 1 - 0 = 1$ (since G is a CDF).

Exercise 6.2.

Exercise 6.3. The proof of Theorem 6.3 follows exactly the same steps (arguments) as the proof of Theorem 6.4 which is presented in Li and Racine (2007, pages 213-214).

Exercise 6.4.

Exercise 6.5. The hints given in Li and Racine (2007, pages 216) provide a detailed solution to this exercise.

Exercise 6.6.

Chapter 7

Semiparametric Partially Linear Models: Solutions

Exercise 7.1. The hint given on Li and Racine (2007, page 246) provides the solution to this problem. Since

$$\hat{\beta}_{inf} = \left[\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\right]^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{Y}_i,$$

where $\tilde{Y}_i = Y_i - E(Y_i|Z_i)$, $\tilde{X}_i = X_i - E(X_i|Z_i)$ and $\tilde{Y}_i = \tilde{X}_i'\beta + u_i$, we have

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{inf} - \beta) &= \sqrt{n}\left(\left[\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\right]^{-1} \sum_{i=1}^{n} \tilde{X}_i(\tilde{X}_i'\beta + u_i) - \beta\right) \\
&= \sqrt{n}\left(\left[\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\right]^{-1} \sum_{i=1}^{n} \tilde{X}_i u_i\right) \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\right]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \tilde{X}_i u_i.
\end{aligned}$$

By the law of large numbers, we have $\frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \xrightarrow{p} E[\tilde{X}_i \tilde{X}_i'] \equiv \Phi = \Phi'$. Also, $E[\tilde{X}_i u_i] = E\{\tilde{X}_i E[u_i|X_i, Z_i]\} = 0$, and $Var[\tilde{X}_i u_i] = E[\tilde{X}_i u_i(\tilde{X}_i u_i)'] = E[u_i^2 \tilde{X}_i \tilde{X}_i'] = E\{E[u_i^2 \tilde{X}_i \tilde{X}_i'|X_i, Z_i]\} = E\{E[u_i^2|X_i, Z_i]\tilde{X}_i \tilde{X}_i'\} = E[\sigma^2(X_i, Z_i)\tilde{X}_i \tilde{X}_i'] \equiv \Psi$.

Using Lindeberg's Central Limit Theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \tilde{X}_i u_i \xrightarrow{d} N(0, \Psi).$$

Hence, we have

$$\begin{aligned}\sqrt{n}(\hat{\beta}_{inf}-\beta) &= \left[\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i\tilde{X}_i'\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_iu_i \\ &\xrightarrow{d} \Phi^{-1}N(0,\Psi) \\ &= N\left(0,\Phi^{-1}\Psi(\Phi^{-1})'\right) \\ &= N\left(0,\Phi^{-1}\Psi\Phi^{-1}\right).\end{aligned}$$

Exercise 7.2.

Exercise 7.3. $\hat{\Phi}_f \xrightarrow{p} \Phi_f$ is proved by Li and Racine (2007, Proposition 7.2, page 243).

$\tilde{u}_i\hat{f}_i = [(Y_i-\hat{Y}_i)-(X_i-\hat{X}_i)'\hat{\beta}]\hat{f}_i$ is an estimator of u_if_i. One can show that $\hat{\Psi}_f = \hat{\Psi}_{f1}+o_p(1)$, where $\hat{\Psi}_{f1} = \frac{1}{n}\sum_{i=1}^{n} u_i^2f_i^2(X_i-\hat{X}_i))(X_i-\hat{X}_i)'\hat{f}_i^2$.

Then by the same steps as in the proof of Proposition 7.2, it can be shown that $\hat{\Psi}_{f1} = \hat{\Psi}_{f2}+o_p(1)$, where $\hat{\Psi}_{f2} = \frac{1}{n}\sum_{i=1}^{n} u_i^2f_i^2V_iV_i'f_i^2$, and $V_i = X_i - E(X_i|Z_i)$.

Hence, $\hat{\Psi} = \hat{\Psi}_{f2}+o_p(1) \xrightarrow{p} \Psi$ by a standard law of large numbers argument.

Exercise 7.4.

Exercise 7.5.

(i) was proved in the proof of Lemma 4 in Li (1996).

(ii) was proved in the proof of Lemma 5 (iii) in Li (1996).

The proofs are quite long and we will not reproduce them here. Readers interested in the proof can consult Li (1996).

Exercise 7.6.

Exercise 7.7.

(i) If $E(u_i^2|X_i,Z_i) = \sigma^2(Z_i)$, we have $E\left(\frac{X_i}{\sigma_i^2}|Z_i\right) = \frac{1}{\sigma_i^2}E(X_i|Z_i)$.

So from (7.32) in Li and Racine (2007), we have

$$\begin{aligned}V_0 &= E\left\{\left[X_i - \frac{E\left(\frac{X_i}{\sigma_i^2}|Z_i\right)}{E\left(\frac{1}{\sigma_i^2}|Z_i\right)}\right]\left[X_i - \frac{E\left(\frac{X_i}{\sigma_i^2}|Z_i\right)}{E\left(\frac{1}{\sigma_i^2}|Z_i\right)}\right]'/\sigma_i^2\right\} \\ &= E\left\{\left[X_i - \frac{\frac{1}{\sigma_i^2}E\left(X_i|Z_i\right)}{\frac{1}{\sigma_i^2}}\right]\left[X_i - \frac{\frac{1}{\sigma_i^2}E\left(X_i|Z_i\right)}{\frac{1}{\sigma_i^2}}\right]'/\sigma_i^2\right\} \\ &= E\left\{[X_i-E(X_i|Z_i)][X_i-E(X_i|Z_i)]'/\sigma_i^2\right\}.\end{aligned}$$

(ii) There is a typo in (7.33) of Li and Racine (2007). The left-hand-side of (7.33) should be $V_{0,R}$, not V_R, so $V_{0,R}$ is defined by (7.33) in Li and Racine (2007). $\sigma^2(Z_i)$ is not defined when $E(u_i^2|X_i,Z_i) \neq E(u_i^2|Z_i)$. However, if we still define $\sigma^2(Z_i)$ by $\sigma^2(Z_i) = E(u_i^2|Z_i)$, we know that $V_0 - V_{0,R}$ is negative semidefinite because V_0 is a semiparametric efficient bound.

Chapter 8

Semiparametric Single Index Models: Solutions

Exercise 8.1.

(i) $E(Y|x) = \sum_{y=0,1} yP(y|x) = (1)P(Y=1|x) + (0)P(y=0|x) = P(Y=1|x)$.

(ii) When $y \in \{1,2\}$ we have $E(Y|x) = \sum_{y=1,2} yP(y|x) = (1)P(Y=1|x) + (2)P(y=2|x) \neq P(Y=1|x)$.

Exercise 8.2.

Exercise 8.3. The purpose of this exercise was to outline a proof for Theorem 8.1. The correct statement should ask students to derive the result of Theorem 8.1 of Li and Racine (2007, page 255) based on (8.1) and (8.2) below.

We assume that $h = cn^{-1/5}$, where c is a positive constant. Define (we omit the trimming functions for notational simplicity)

$$\hat{S}(\theta) = \sum_{i=1}^{n} [Y_i - \hat{E}(Y_i|x_i'\beta)]^2, \tag{8.1}$$

where $\hat{E}(Y_i|x_i'\beta) = (nh)^{-1}\sum_{j\neq i} Y_j K_{ij,\beta}/\hat{p}(x_i\beta)$, $\hat{p}(x_i'\beta) = (nh)^{-1}\sum_{j\neq i} K_{ij,\beta}$.

Then Ichimura (1993) and Härdle, Hall and Ichimura (1993) have proved the following:

$$\hat{S}(\beta) = \tilde{S}(\beta) + (s.o.), \tag{8.2}$$

where $(s.o.)$ contains terms unrelated to β (they may be related to β_0) and terms that are of smaller orders than $\tilde{S}(\beta)$, and

$$\tilde{S}(\beta) = \sum_{i=1}^{n} \{Y_i - E[g(x_i'\beta_0)|x_i'\beta]\}^2. \tag{8.3}$$

Noting that (8.3) is a parametric model, minimizing (8.3) with respect to β leads to a NLS estimator of β. Applying a Taylor expansion to $g(x_i'\beta_0)$ and $E[g(x_i'\beta_0)|x_i'\beta]$ at β we obtain

$$\begin{aligned} g(x_i'\beta_0) &= g(x_i'\beta) + g^{(1)}(x_i'\tilde{\beta})x_i'(\beta_0 - \beta) \\ &= g(x_i'\beta) + g^{(1)}(x_i'\beta_0)x_i'(\beta_0 - \beta) + (s.o.) \qquad (8.4) \\ E[g(x_i'\beta_0)|x_i'\beta] &= g(x_i'\beta) + E[g^{(1)}(x_i'\tilde{\beta})x_i'(\beta_0 - \beta)|x_i'\beta] \\ &= g(x_i'\beta) + g^{(1)}(x_i'\beta_0)E[x_i|x_i'\beta_0]'(\beta_0 - \beta) + (s.o.) \qquad (8.5) \end{aligned}$$

From (8.3), (8.4), (8.5) and using $Y_i = g(x_i'\beta_0) + u_i$, we obtain

$$\begin{aligned} \tilde{S}(\beta) &= \sum_{i=1}^{n} \left\{ u_i - g^{(1)}(x_i'\beta_0) \left[x_i - E(x_i|x_i'\beta_0) \right]' (\beta - \beta_0) \right\}^2 + (s.o.) \\ &\equiv \sum_{i=1}^{n} \left[u_i - z_i'(\beta - \beta_0) \right]^2 + (s.o.), \end{aligned}$$

where

$$z_i = g^{(1)}(x_i'\beta_0)[x_i - E(x_i|x_i'\beta_0)].$$

Minimizing the leading term of $\tilde{S}(\beta)$ with respect to β gives $\hat{\beta} - \beta_0 = [\sum_i z_i z_i']^{-1} \sum_i z_i u_i$. Hence, we have

$$\sqrt{n}(\hat{\beta} - \beta_0) = \left[\frac{1}{n} \sum_{i=1}^{n} z_i z_i' \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i}^{n} z_i u_i \xrightarrow{d} N(0, \Omega), \qquad (8.6)$$

where $\Omega = V^{-1}\Sigma V^{-1}$ with $V = E[z_i z_i']$ and $\Sigma = E[\sigma^2(x_i) z_i z_i']$, which is the same as Theorem 8.1 of Li and Racine (2007, page 255) except that we omit the trimming function.

To complete the proof for Theorem 8.1, one also needs to establish (8.2). The proof of (8.2) can be found in Ichimura (1993) and Härdle et al. (1993). Since the proof is quite long we will not reproduce it here.

Exercise 8.4.

Chapter 9

Additive and Smooth (Varying) Coefficient Semiparametric Models: Solutions

Exercise 9.1. There are some typos in the expression of this exercise. The correction expression should be:

"Let $\hat{g}(z)$ be defined as in (9.25) show that

$$\sqrt{nh}\left[\hat{g}(z) - g(z) - \nu - \frac{\kappa_2}{2}\mu(z)\right] \xrightarrow{d} N\left(0, \sum_{\alpha=1}^{q} V_\alpha(z_\alpha)\right),$$

where $\nu = \sum_{\alpha=1}^{q} \nu_\alpha$, $\mu = \sum_{\alpha=1}^{q} \mu_\alpha$, $\nu_\alpha = \nu_\alpha(z)$ and $\mu_\alpha = \mu_\alpha(z_\alpha)$ are defined in Li and Racine (2007, page 294)."

Letting $c_0 = E(Y_i)$ then we know that $\bar{Y} - c_0 = O_p(n^{-1/2})$. Note that $g(z) = C_0 + \sum_{\alpha=1}^{q} g_\alpha(z_\alpha)$ and $\bar{Y} - C_0 = O_p\left(n^{-1/2}\right)$. Then by (9.25) and using (9.33) we have

$$\begin{aligned}
&\sqrt{nh}\left\{\hat{g}(z) - g(z) - \sum_{\alpha=1}^{q}\nu_\alpha - (\kappa_2/2)h^2\sum_{\alpha=1}^{q}\mu_\alpha\right\} \\
&= \sqrt{nh}\left\{\bar{Y} - c_0 + \sum_{\alpha=1}^{q}(\hat{g}_\alpha(z_\alpha) - g_\alpha(z_\alpha)) - \sum_{\alpha=1}^{q}\nu_\alpha - (\kappa_2/2)h^2\sum_{\alpha=1}^{q}\mu_\alpha\right\} \\
&= \sqrt{nh}\left\{\sum_{\alpha=1}^{q}\left[\hat{g}_\alpha(z_\alpha) - g_\alpha(z_\alpha) - \sum_{\alpha=1}^{q}\nu_\alpha - (\kappa_2/2)h^2\sum_{\alpha=1}^{q}\mu_\alpha\right] + O_p(\sqrt{h})\right\} \\
&\xrightarrow{d} N\left(0, \sum_{s=1}^{q} V_\alpha(z)\right)
\end{aligned}$$

by (9.33) and the fact that $\hat{g}_\alpha(z_\alpha)$, $\alpha = 1, \ldots, q$, are asymptotically independent of each other (their asymptotic covariances are zero by (9.33)).

Exercise 9.2.

Exercise 9.3. Note that $\zeta_i = W_i - X_i'\beta_w(Z_i)$ can be estimated by $\hat{\zeta}_i = W_i - X_i'\hat{\beta}_w(Z_i)$. Hence, consistent estimators of A and B are given by

$$\hat{A} = n^{-1}\sum_{i=1}^{n}\hat{\zeta}_i\hat{\zeta}_i' \text{ and } \hat{B} = n^{-1}\sum_{i=1}^{n}\hat{\zeta}_i\hat{\zeta}_i'\hat{u}_i^2,$$

where $\hat{u}_i = Y_i - W_i'\hat{\gamma} - X_i'\hat{\beta}(Z_i)$, $\hat{\beta}(Z_i)$ is defined in (9.69) with γ replaced by $\hat{\gamma}$, $\hat{\gamma}$ is the OLS estimator of γ based on (9.70).

Exercise 9.4.

Chapter 10

Selectivity Models: Solutions

Exercise 10.1. Using the result that if $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$, then

$$E(v_2|v_1 > c) = \mu_2 + \rho\sigma_{22}^{1/2}\frac{\phi(\alpha)}{\Phi(\alpha)}, \tag{10.1}$$

where $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$, $\alpha = (c-\mu_1)/\sigma_{11}^{1/2}$, $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of a standard normal random variable.

Using (10.1) we have

$$\begin{aligned} E(u_{2i}|X_i, Y_{1i} = 1) &= E(u_{2i}|X_i, u_{1i} > -X_{1i}'\beta_1) = E(u_{2i}|u_{1i} > -X_{1i}'\beta_1) \\ &= \sqrt{\sigma_{22}}\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}\frac{\phi(-X_{1i}'\beta_1/\sigma_{11}^{1/2})}{\Phi(X_{1i}'\beta_1/\sigma_{11}^{1/2})} \\ &= \sigma_{11}^{-1/2}\sigma_{12}\frac{\phi(X_{1i}'\beta_1/\sigma_{11}^{1/2})}{\Phi(X_{1i}'\beta/\sigma_{11}^{1/2})}. \end{aligned}$$

The last equality follows because $\phi(v) = \phi(-v)$.

Exercise 10.2.

Exercise 10.3.

(i) Under H_0^a, $Y_{2i} = X_{2i}'\beta + u_{2i}$ with $E(u_{2i}|X_i, Y_{2i} > 0) = 0$, we have

$$\hat{u}_{2i} = Y_{2i} - X_{2i}'\hat{\beta}_2 = X_{2i}'(\beta_2 - \hat{\beta}_2) + u_{2i} = u_{2i} + O_p(n^{-1/2}),$$

and

$$\hat{u}_{1i} = Y_{1i} - X_{1i}'\hat{\beta}_1 = X_{1i}'(\beta_1 - \hat{\beta}_1) + u_{1i} = u_{1i} + O_p(n^{-1/2}),$$

one can show that

$$\tilde{I}_{1n}^a = \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} u_{2i}u_{2j}K_h(u_{1i} - u_{2i}) + (s.o.) \equiv I_{n,0}^a + (s.o.).$$

It is obvious that $E(I^a_{n,0}) = 0$ because $E(u_{2i}|u_{1i}) = 0$ under H^a_0. It is straightforward to show that $Var(I^a_{n1}) = E[(I^a_{n,0})^2] = O((n^2h)^{-1}) = o(1)$ (by a derivation similar to that in the proof of Theorem 12.1 of Li and Racine (2007)).

Hence, $I^a_{n,0} = o_p(1)$ which in turn implies that $\tilde{I}^a_n = I^a_{n,0} + (s.o.) = o_p(1)$ under H^a_0.

(ii) Under H^a_1, $\hat{u}_{2i} = Y_{2i} - X'_{2i}\hat{\beta}_2 = X'_{2i}(\beta_2 - \hat{\beta}_2) + g(u_{1i}) + u_{2i} = g(u_{1i}) + u_{2i} + O_p(n^{-1/2})$, where $g(u_{1i}) = E(u_{2i}|u_{1i})$. Then it can be shown that

$$
\begin{aligned}
\tilde{I}^a_n &= \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1}[g(u_{1i}) + u_{2i}][g(u_{1j}) + u_{2j}]K_h(u_{1i} - u_{2i}) + (s.o.) \\
&= \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1} g(u_{1i})g(u_{1j})K_h(u_{1i} - u_{2i}) + (s.o.) \\
&\equiv I^a_{n,1} + (s.o.).
\end{aligned}
$$

Now,

$$
\begin{aligned}
E[I^a_{n,1}] &= E[g(u_{1i})g(u_{1j})K_h(u_{1i} - u_{2i})] \\
&= \int\int f(u_{1i})f(u_{1j})g(u_{1i})g(u_{1j})K_h(u_{1i} - u_{2i})du_{1i}du_{1j} \\
&= \int\int f(u_{1i})f(u_{1i} + hv)g(u_{1i})g(u_{1i} + hv)K(v)du_{1i}dv \\
&= \left[\int f(u_{1i})f(u_{1i}g(u_{1i})g(u_{1i})du_{1i}\right]\left[\int K(v)dv\right] + O(h^2) \\
&= E\left\{f(u_{1i})[g(u_{1i})]^2\right\} + o(1) \equiv C + o(1),
\end{aligned}
$$

where $C = E\left\{f(u_{1i})[g(u_{1i})]^2\right\} > 0$ under H^a_1.

It is straightforward to show that $Var(I^a_{n,1}) = o(1)$. Hence, $\tilde{I}^a_n = I^2_{n,1} + (s.o.) = C + o_p(1)$.

(iii) Under H^a_1, $\tilde{I}^a_n = C + o_p(1)$. Also, it is easy to see that $\hat{\sigma}^2_a$ is $O_p(1)$ under either H^a_0 or H^a_1 ($\hat{\sigma}_a$ is defined in Li and Racine (2007, Proposition 10.1)). Hence, $T^a_n \stackrel{def}{=} nh^{1/2}\tilde{I}^a_n/\hat{\sigma}_a = nh^{1/2}C + (s.o.)$ which diverges to $+\infty$ at the rate of $nh^{1/2}$. The test statistic has a standard normal distribution under H^a_0. If we choose a 5% level test, the (one-sided) critical value is 1.645, hence we have

$$
P(\text{reject } H^a_0|H^a_0 \text{ is false}) = P(T_n > 1.645) \to 1 \text{ as } n \to \infty
$$

because $T_n \to +\infty$. Hence, the test statistic T^a_n has asymptotic power equal to one.

Chapter 11

Censored Models: Solutions

Exercise 11.1. Letting $x \sim N(\mu, \sigma^2)$, then the density for truncated variable $x|(x > c)$ is

$$\frac{f(x)}{P[x > c]} = \frac{f(x)}{[1 - \Phi(\alpha)]},$$

where $\alpha = (c - \mu)/\sigma$ because

$$P(x > c) = 1 - P(x \leq c) = 1 - P\left(\frac{x - \mu}{\sigma} \leq \frac{c - \mu}{\sigma}\right) = 1 - \Phi(\alpha).$$

Hence, one can show that

$$\begin{aligned} E[x|x > c] &= \frac{\int_c^\infty x f(x) dx}{[1 - \Phi(\alpha)]} = \mu + \sigma \frac{\phi(\alpha)}{[1 - \Phi(\alpha)]} \\ &= \mu + \sigma \frac{\phi(-\alpha)}{\Phi(-\alpha)} \end{aligned}$$

because $\phi(-a) = \phi(a)$ and $1 - \Phi(a) = \Phi(-a)$.

Applying the above result to $x = \epsilon_i \sim N(0, \sigma^2)$, we have $\mu = 0$ and $c = -X_i'\beta$. Hence we obtain

$$E[\epsilon_i|\epsilon_i > -X_i'\beta] = \sigma \frac{\phi(X_i'\beta)}{\Phi(X_i'\beta)}.$$

Exercise 11.2.

Exercise 11.3. *Proof.* We have that

$$\begin{aligned} \hat{F}_n(y) &= 1 - \prod_{i=1}^{n} \left[1 - \frac{\delta_{[i:n]}}{n - i + 1}\right]^{\mathbf{1}(Z_{i:n} \leq y)}, \\ \tilde{F}_n(y) &= \sum_{i=1}^{n} W_{in} \mathbf{1}(Z_{i:n} \leq y), \end{aligned}$$

where

$$W_{in} = \frac{\delta_{[i:n]}}{n - i + 1} \prod_{j=1}^{i-1} \left[\frac{n - j}{n - j + 1}\right]^{\delta_{[j:n]}} = \frac{\delta_{[i:n]}}{n - i + 1} \prod_{j=1}^{i-1} \left[1 - \frac{1}{n - j + 1}\right]^{\delta_{[j:n]}}.$$

It is easy to see that

$$\left[1 - \frac{1}{n-j+1}\right]^{\delta_{[j:n]}} = \left[1 - \frac{\delta_{[j:n]}}{n-j+1}\right], \text{ for any } j.$$

For any $y \in \mathcal{R}$ there exists an integer $m \geq 1$ such that $Z_{m:n} \leq y < Z_{(m+1):n}$. Then we have that

$$\begin{aligned}
\hat{F}_n(y) &= 1 - \prod_{i=1}^{m}\left[1 - \frac{\delta_{[i:n]}}{n-i+1}\right], \\
\tilde{F}_n(y) &= \sum_{i=1}^{m} \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1}\left[1 - \frac{\delta_{[j:n]}}{n-j+1}\right] \\
&= \frac{\delta_{[1:n]}}{n-1+1} + \frac{\delta_{[2:n]}}{n-2+1}\left[1 - \frac{\delta_{[1:n]}}{n-1+1}\right] + \frac{\delta_{[3:n]}}{n-3+1}\left[1 - \frac{\delta_{[1:n]}}{n-1+1}\right]\left[1 - \frac{\delta_{[2:n]}}{n-2+1}\right] + \cdots \\
&\quad + \frac{\delta_{[m:n]}}{n-m+1}\prod_{j=1}^{m-1}\left[1 - \frac{\delta_{[j:n]}}{n-j+1}\right] \\
&= 1 - \prod_{i=1}^{m}\left[1 - \frac{\delta_{[i:n]}}{n-i+1}\right].
\end{aligned}$$

So $\hat{F}_n(y) \equiv \tilde{F}_n(y)$, which completes the proof. □

Exercise 11.4.

Chapter 12

Model Specification Tests: Solutions

Exercise 12.1. Letting $\mu_j = E(x_i^j)$ then we have $\mu_1 = \mu_3 = 0$. Note that $E(y_i) = E(\alpha + \beta_1 x_i + \beta_3 x_i^3 + u_i) = \alpha$, $E(x_i y_i) = E[x_i(\alpha + \beta_1 x_i + \beta_3 x_i^3 + u_i)] = \beta_1\mu_2 + \beta_3\mu_4$, $E(x_i^2 y_i) = \alpha\mu_2$.

(i) The least squares estimator based on model (12.3) is

$$
\begin{aligned}
\begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= (X'X)^{-1}XY = (X'X/n)^{-1}(XY)/n \\
&= \begin{pmatrix} 1 & \frac{1}{n}\sum_i x_i & \frac{1}{n}\sum_i x_i^2 \\ \frac{1}{n}\sum_i x_i & \frac{1}{n}\sum_i x_i^2 & \frac{1}{n}\sum_i x_i^3 \\ \frac{1}{n}\sum_i x_i^2 & \frac{1}{n}\sum_i x_i^3 & \frac{1}{n}\sum_i x_i^4 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{n}\sum_i y_i \\ \frac{1}{n}\sum_i x_i y_i \\ \frac{1}{n}\sum_i x_i^2 y_i \end{pmatrix} \\
&\xrightarrow{p} \begin{pmatrix} 1 & 0 & \mu_2 \\ 0 & \mu_2 & 0 \\ \mu_2 & 0 & \mu_4 \end{pmatrix}^{-1} \begin{pmatrix} \alpha \\ \beta_1\mu_2 + \beta_3\mu_4 \\ \alpha\mu_2 \end{pmatrix} \\
&= \frac{1}{\mu_2(\mu_4 - \mu_2^2)} \begin{pmatrix} \mu_2\mu_4 & 0 & -\mu_2 \\ 0 & \mu_4 - \mu_2^2 & 0 \\ -\mu_2 & 0 & \mu_2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1\mu_2 + \beta_3\mu_4 \\ \alpha\mu_2 \end{pmatrix},
\end{aligned}
$$

from which we obtain $\hat{\beta}_2 \xrightarrow{p} (\alpha\mu_2 - \alpha\mu_2)/[\mu_2(\mu_4 - \mu_2^2)] = 0$.

(ii) It is easy to show that $\sqrt{n}\hat{\beta}_2 \to N(0, V_2)$, where $V_2 = \text{plim}[n\,Var(\hat{\beta}_2)]$ is a finite positive constant. Hence, the t-statistic: $t_{\hat{\beta}_2} = \hat{\beta}_2/\sqrt{Var(\hat{\beta}_2)} = \sqrt{n}\hat{\beta}_2/\sqrt{n\,Var(\hat{\beta}_2)} \xrightarrow{d} N(0,1)$. Therefore, its asymptotic power equals the size of the test.

Exercise 12.2.

Exercise 12.3.

(i) Obviously $E(I_{2n}) = 0$ and

$$\begin{aligned}
E\left[||I_{2n}||^2\right] &= \frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i}\sum_{i'}\sum_{j'\neq i'} E[u_i u_{i'} Z_j' Z_{j'} K_{h,ij} K_{h,i'j'}] \\
&= \frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i}\sum_{j'} E[u_i^2 Z_j' Z_{j'} K_{h,ij} K_{h,ij'}] \\
&= \frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i} E[u_i^2 Z_j' Z_j K_{h,ij}^2] \\
&\quad + \frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i}\sum_{j'\neq i, j'\neq j} E[u_i^2 Z_j' Z_{j'} K_{h,ij} K_{h,ij'}] \\
&= n^{-1}E[\sigma^2(x_1) Z_2' Z_2 K_{h,ij}] + \frac{n-1}{n-2}E[\sigma^2(x_1) Z_2' Z_3 K_{h,12} K_{h,13}] \\
&= n^{-1}O\left((h_1\dots h_q)^{-1}\right) + O(1) = O(1).
\end{aligned}$$

Hence, $I_{2n} = O_p(1)$.

$$\begin{aligned}
E\left[|I_{3n,ts}|\right] &= E\left[\left|\frac{1}{n(n-1)}\sum_i\sum_{j\neq i} Z_{it} Z_{is} K_{h,ij}\right|\right] \\
&\leq \frac{1}{n(n-1)}\sum_i\sum_{j\neq i} E[|Z_{it} Z_{is} K_{h,ij}|] = E[|Z_{1t} Z_{2s} K_{h,12}|] = O(1).
\end{aligned}$$

Hence, $I_{3n} = O(1)$.

(ii) Under H_0^a, $\hat{u}_i = u_i + X_i'(\beta - \hat{\beta}) = u_i + O_p\left(n^{-1/2}\right)$, and $\hat{u}_i^2 = u_i^2 + O_p\left(n^{-1/2}\right)$. Hence,

$$\begin{aligned}
\hat{\sigma}_a^2 &= \frac{2h_1\dots h_q}{n(n-1)}\sum_i\sum_{j\neq i}\hat{u}_i^2\hat{u}_j^2 K_{h,ij}^2 \\
&= \frac{2h_1\dots h_q}{n(n-1)}\sum_i\sum_{j\neq i} u_i^2 u_j^2 K_{h,ij}^2 + (s.o.) = \sigma_{a,0}^2 + (s.o.).
\end{aligned}$$

$$\begin{aligned}
E[\sigma_{a,0}^2] &= 2(h_1\dots h_q)E\left[u_i^2 u_j^2 K_{h,ij}^2\right] \\
&= 2(h_1\dots h_q)E[\sigma^2(x_i)\sigma^2(x_j)K_{h,ij}^2] \\
&= 2\int \sigma^4(x_i) f(x_i)^2 dx_i + (s.o.) \\
&= 2E\left[\sigma^4(X_i) f(X_i)\right] + (s.o.) \\
&\equiv \sigma_a^2 + o(1).
\end{aligned}$$

It is straightforward to show that $Var(\sigma_{a,0}^2) = o(1)$. Hence, $\hat{\sigma}_a^2 = \sigma_{a,0}^2 + (s.o.) = \sigma_a^2 + o_p(1)$.

(iii) Obviously $E(I_{1n}) = 0$ and by following the same arguments one can show that $Var(I_{1n}) = (n^2H_q)^{-1}\left(\sigma_a^2 + o(1)\right)$, where $H_q = h_1 \dots h_q$. Hence, $nH_q^{1/2}I_{1n} \sim \left(0, \sigma_a^2\right)$.

Now let $H_{n,ij} = u_i u_j K(X_i - X_j)/h)$, $W_i = (X_i, u_i)$, $G_{n,ij} = E[H_{n,il}H_{n,jl}|W_i, W_j]$. Then it is easy to show that $E[G_{n,ij}] = O(H_q^3)$, $E[H_{n,ij}^4] = O(H_q)$ and $E[H_{n,ij}^2] = O(H_q)$. Hence,

$$\frac{E[G_{n,ij}] + n^{-1}E[H_{n,ij}^4]}{[E(H_{n,ij}^2)]^2} = O\left(H_q + (nH_q)^{-1}\right) = o(1).$$

Hence, the condition for Hall's (1984) Central Limit Theorem holds and we have

$$n\sqrt{H_q}I_{1n} \xrightarrow{d} N(0, \sigma_a^2)$$

(iv) By (i) - (iii) we have under H_0^a that

$$nH_q^{1/2}I_n^a/\sqrt{\hat{\sigma}_a^2} = nH_q^{1/2}I_{1n}/\sqrt{\sigma_a^2} + (s.o.) = nH_q^{1/2}I_{1n}/\sqrt{\sigma_a^2} + (s.o.) \xrightarrow{d} N(0,1).$$

Exercise 12.4.

Exercise 12.5. We will consider the case of a linear model: $Y = X\gamma + u$. Since $Y^* = X\hat{\gamma} + u^*$, we have

$$\begin{aligned}\hat{\gamma}^* &= (X'X)^{-1}X'Y^* = \hat{\gamma} + (X'X)^{-1}X'u^* \\ &= \hat{\gamma} + (X'X/n)^{-1}X'u^*/n = \hat{\gamma} + O_p\left(n^{-1/2}\right)\end{aligned}$$

since $X'X/n = E(X_iX_i') + o_p(1)$ and $X'u^*/n = O_p^*\left(n^{-1/2}\right)$. This is because $E^*(X'u^*/n) = n^{-1}\sum_i X_i E^*(u_i^*) = 0$ and $E^*\left[||(X'u^*/n)||^2\right] = n^{-2}\sum_i\sum_j X_i'X_jE^*(u_i^*u_j^*) = n^{-2}\sum_i X_i'X_i\hat{u}_i^2 = O_p\left(n^{-1}\right)$.

Exercise 12.6.

Exercise 12.7. Define $H_{n,ij} = 2(K_{h,ij}^x)^2(K_{h,ij}^y)^2$ and note that $\hat{\sigma}_f^2 = \frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j>i}^n H_{n,ij}$ is a second order U-statistic. By the H-decomposition we have

$$\begin{aligned}\hat{\sigma}_f^2 &= E[H_{n,ij}^2] + o_p(1) = 2E[K_{h_x}(X_i - X_j)]E[K_{h_y}(Y_i - Y_j)] \\ &= 2\kappa^p\left\{E[f_1(X_i)^2] + o(1)\right\}\kappa^q\left\{E[f_2(Y_i)^2] + o(1)\right\} \\ &= \sigma_f^2 + o_p(1).\end{aligned}$$

Using Lemma A.16 it is easy to show that $Var(\hat{\sigma}_f^2) = o(1)$. Hence, $\hat{\sigma}_f^2 = \sigma_f^2 + o_p(1)$.

Chapter 13

Nonsmoothing Tests: Solutions

Exercise 13.1.

$$
\begin{aligned}
E\left[||Z_i(.)||_\nu^2\right] &= E\left[\int [Z_i(x)]^2 d\nu(x)\right] \\
&= E\left\{\int \left[\mathcal{H}(X_i,x) + \phi(X_i,x)\right]^2 u_i^2 dx\right\} \\
&= E\left\{\int \left[\mathcal{H}(X_i,x) + \phi(X_i,x)\right]^2 \sigma^2(X_i) d\nu(x)\right\} \\
&\le CE[\sigma^2(X_i)]
\end{aligned}
$$

provided that $|\mathcal{H}(x_i,x) + \phi(x_i,x)|$ is a bounded function and that $\int d\nu(x)$ is finite. If ν is the Lebesgue measure ($d\nu(x) = dx$), then we require that the integration set over x is a bounded set.

This does not necessarily restrict X_i to take values in a bounded set because, even if X_i's support is unbounded, one can map X_i into a compact (hence, bounded) set, say $\mathcal{H}(X_i,x)$, provided that $\mathcal{H}(X_i,x)$ and X_i generates the same σ-algebra.

Exercise 13.2.

Chapter 14

K-Nearest Neighbor Methods: Solutions

We first make some comments on how to work out the problems in Chapter 14.

(i) If we have a term like $A(X_i, x) = R_x^{-q} g(x_i) w((x_i - x)/R_x)$, where x_i is random and x is a given (fixed and non-random) point, we can use Lemma 14.1 to evaluate such a term easily. This is because we only need to conditional on x_i to compute terms like $E[R_x^{-q} g(x_i) w((x_i - x)/R_x)] = E\{R_x^{-q} E[g(x_i) w((x_i - x)/R_x)|x_i]\}$. Lemma 14.1 can be used to readily yield this result. This is the case for exercises 14.1 and 14.2.

(ii) If we have a term like $R_i^{-q} g(z_i, z_j) w((z_i - z_x)/R_i)$, where both z_i and z_j are random variables, then Lemma 14.1 is not convenient to use. In this case we first use (14.32) because in this way we first condition on (x_i, R_i) and integrate out x_j. Then we can use Lemma 14.1. This is the case for exercises 14.3 to 14.5.

Exercise 14.1. Note that there is a typo in the expression of this exercise. It should be $A_{21}^{1,x} = o(1)$, not $A_{21}^{1,x} = O(n^{-1})$.

$A_{21}^{1,x} = \frac{1}{nR_x^{q+2}} \sum_{i=1}^n w_{i,x}(x_i - x)$, where $w_{i,x} = w((X_i - x)/R_x)$, is defined in (14.37) in Li and Racine (2007, page 436).

$$
\begin{aligned}
Var(A_{21}^{1,x}) &= \frac{1}{n^2} \sum_{i=1}^n Var\left[R_x^{-(q+2)} w_{i,x}(x_i - x)\right] \le \frac{1}{n^2} \sum_{i=1}^n E\left[R_x^{-2(q+2)} w_{i,x}^2 (x_i - x)^2\right] \\
&= \frac{1}{n} E\left\{E\left[R_x^{-2(q+2)} w_{i,x}^2 (x_i - x)(x_i - x)' | R_x\right]\right\} \\
&= \frac{1}{n} E\left[f(x) R_x^{-(q+2)} \int w(v)^2 vv' dv\right] + (s.o.) \\
&= \frac{\kappa_2}{n} f(x) E[R_x^{-(q+2)}] + (s.o.) \\
&= \kappa_2 n^{-1} O\left((k/n)^{-(q+2)/q}\right) = O\left((n^2/k^{q+2})^{1/q}\right) = o(1)
\end{aligned}
$$

by Lemma 14.1 and Assumption 14.4 that $n^2/k^{q+2} = o(1)$ (Li and Racine (2007, page 421)).

Exercise 14.2.

Exercise 14.3. We have

$$\begin{aligned}
S_{g-\hat{g}} &= \frac{1}{n}\sum_{i=1}^{n}(g_i-\hat{g}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j\neq i}^{n}(g_j-g_i)R_i^{-q}w_{ij}\right]^2 (I_1/\hat{f}_1)^2 \\
&\leq b^{-2}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j\neq i}^{n}(g_j-g_i)R_i^{-q}w_{ij}\right]^2 \\
&= \frac{1}{n^3b^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}R_i^{-2q}(g_j-g_i)^2w_{ij}^2 + \frac{1}{n^3b^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l\neq i,j}^{n}R_i^{-2q}(g_j-g_i)w_{ij}(g_l-g_i)w_{il} \\
&= S_1+S_2.
\end{aligned}$$

$$\begin{aligned}
E[|S_1|] &= \frac{1}{nb^2}E\left[R_1^{-2q}(g_1-g_2)^2w_{12}^2\right] \\
&= \frac{1}{nb^2}E\left\{E\left[R_1^{-2q}w_{12}^2(g_1-g_2)^2|z_1,R_1\right]\right\} \\
&= \frac{k}{n^2b^2}E\left[\frac{1}{R_1^{2q}G(R_1)}\int f(z_2)w_{12}^2(g_1-g_2)^2dz_2\right] \qquad \text{by (14.32)} \\
&= \frac{k}{n^2b^2}E\left[\frac{1}{R_1^{q}G(R_1)}f(z_1)R_1g^{(1)}(z_1)'\int w(v)^2vv'dvg^{(1)}(z_1)R_1\right] + (s.o.) \\
&= \frac{k}{n^2b^2}O(1)E\left[\frac{1}{R_1^{q-2}G(R_1)}|z_1\right] + (s.o.) \\
&= \frac{k}{n^2b^2}O\left((k/n)^{2/q-2}\right) = O\left((k/n)^{2/q}k^{-1}b^{-2}\right)
\end{aligned}$$

by Lemma 14.1.

$$
\begin{aligned}
E|S_2| &= b^{-2}\left|E\left[R_1^{-2q}(g_1-g_2)(g_1-g_3)w_{12}w_{13}\right]\right| \\
&= b^{-2}\left|E\left\{R_1^{-q}E\left[(g_1-g_2)w_{12}|z_1,R_1\right]E\left[R_1^{-q}(g_1-g_3)w_{13}|z_1,R_1\right]\right\}\right| \\
&= \frac{1}{b^2}E\left\{\left[R_1^{-q}\left(\frac{k}{n}\frac{1}{G(R_1)}\int f(z_2)[g_2-g_1]w_{12}dz_2\right)\right]^2\right\} \\
&= \frac{k^2}{n^2b^2}E\left\{\left[\frac{1}{G(R_1)}\int f(z_1+R_1v)[g(z_1+R_1v)-g_1]w(v)dv\right]^2\right\} \\
&= \frac{k^2}{n^2b^2}O(1)E\left\{\left[\frac{R_1^{\nu}}{G(R_1)}\right]^2\right\} \\
&= \frac{k^2}{n^2b^2}O\left(E\left[\frac{R_1^{(2\nu)}}{G(R_1)^2}\right]\right) \\
&= \frac{k^2}{n^2b^2}O\left((k/n)^{2\nu/q-2}\right) = O\left((k/n)^{2\nu/q}b^{-2}\right)
\end{aligned}
$$

by Lemma 14.1.

Exercise 14.4.

Exercise 14.5. This is Lemma 8 of Liu and Lu (1997). We will not reproduce it here.

Chapter 15

Nonparametric Series Methods: Solutions

Exercise 15.1. Under the conditional homoskedastic error assumption, we can estimate Σ by

$$\hat{\Sigma} = \hat{\sigma}^2 \left[\frac{1}{n} \sum_i p^K(x_i) p^K(x_i)' \right] = \hat{\sigma}^2 \hat{Q},$$

where $\hat{\sigma}^2 = n^{-1} \sum_i \hat{u}_i^2$.

Now we have

$$\begin{aligned} \hat{V}_K &= p^K(x) \hat{Q}^{-1} \hat{\Sigma} \hat{Q}^{-1} p^K(x) = \hat{\sigma}^2 p^K(x) \hat{Q}^{-1} p^K(x) \\ &= \hat{\sigma}^2 p^K(x) Q^{-1} p^K(x) + (s.o.) = \sigma^2 p^K(x) Q^{-1} p^K(x) + (s.o.) \\ &= V_K + (s.o.), \end{aligned}$$

where we have used Lemma 15.2 $\hat{Q} = Q + O_p(\zeta_0(K)K/n)$, where $\zeta_0(K) = O(K)$ for power series and $\zeta_0(K) = O(\sqrt{K})$ for splines (e.g., Newey (1997)) and $\hat{\sigma}^2 = n^{-1} \sum_i u_i^2 + o_p(1) = \sigma^2 + o_p(1)$. Hence,

$$n^{1/2} \hat{V}_k^{-1/2} (\hat{g}(x) - g(x)) = n^{1/2} V_k^{-1/2} (\hat{g}(x) - g(x)) + (s.o.) = A_n + (s.o.) \xrightarrow{d} N(0,1)$$

follows the arguments given in the hint to Exercise 15.1 in Li and Racine (2007).

Exercise 15.2.

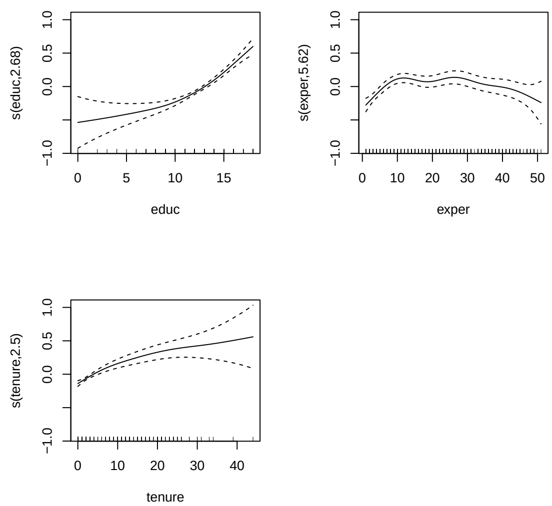
Exercise 15.3. Write a program for additive models using a power series estimator and use Wang and Yang's (2005) BIC criterion to select the significant variables.

See the exercise below.

Exercise 15.4. Repeat Exercise 15.3 but use the leave-one-out method to select the number of series terms.

Here we conduct a similar exercise using the mgvc library in R. We shall use the wage1 data by way of example. See ?gam for details. GACV.cp uses a Cp/UBRE/AIC criterion.

```
R> library(mgcv)
R> data(wage1)
R> attach(wage1)
R> model.gam <- gam(lwage~s(educ)+s(exper)+s(tenure),method="GACV.Cp")
R> plot(model.gam,pages=1)
```



Exercise 15.5. Multiplying both sides of (15.35) by $p_j(z_\alpha)$ and integrating over z_α we obtain

$$\int g_\alpha(z_\alpha)p_j(z_\alpha)dz_\alpha = \sum_{l=1}^{\infty}\theta_{\alpha l}\int p_l(z_\alpha)p_j(z_\alpha)dz_\alpha = \sum_{l=1}^{\infty}\theta_{\alpha l}\delta_{lj} = \theta_{\alpha j},$$

where $\delta_{jl} = 1$ if $j = l$ and 0 otherwise (the Kronecker delta function).

Exercise 15.6.

Chapter 16

Instrumental Variables and Efficient Estimation of Semiparametric Models: Solutions

Exercise 16.1. This exercise requires the assumptions that $E(S_{t-1}) = 0$, $E(Z_{t-1}^3) = 0$ and that Z_{t-1} is independent of Y_{t-2}. With $\theta(Z_t) = Z_t^2$ we have $Y_{t-1} = Y_{t-2}\beta + Z_{t-1}^2 + u_{t-1}$ so that

$$E(Y_{t-1}Z_{t-1}) = E(Y_{t-2}Z_{t-1})\beta + E(Z_{t-1}^3) + E(u_{t-1}Z_{t-1}) = 0$$

since we have assumed that $E(Z_{t-1}) = 0$, $E(Z_{t-1}^3) = 0$, Z_{t-1} is independent of Y_{t-2} and u_{t-1} (for example, if Z_t is an i.i.d. sequence).

Exercise 16.2.

Exercise 16.3. *Proof.* From the definition, we have

$$\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha_0] = E\left\{\frac{d\rho(Z,\alpha_0)}{d\alpha}[\alpha - \alpha_0]|X\right\}, \text{ for all } \alpha \in \mathcal{A}.$$

By Jensen's inequality and the law of iterated expectations, it is easy to obtain that

$$\begin{aligned}
\|\alpha - \alpha_0\|_w &= \sqrt{E\left\{\left[\frac{dm(X,\alpha_0)}{d\alpha}(\alpha - \alpha_0)\right]^2/\sigma^2(X)\right\}} \\
&= \sqrt{E\left\{\left(E\left[\left(\frac{d\rho(Z,\alpha_0)}{d\alpha}(\alpha - \alpha_0)\right)|X\right]\right)^2/\sigma^2(X)\right\}} \\
&\leq \sqrt{E\left\{E\left[\left(\frac{d\rho(Z,\alpha_0)}{d\alpha}(\alpha - \alpha_0)\right)^2|X\right]/\sigma^2(X)\right\}} \\
&= \sqrt{E\left\{\left[\frac{d\rho(Z,\alpha_0)}{d\alpha}(\alpha - \alpha_0)\right]^2/\sigma^2(X)\right\}} = \|\alpha - \alpha_0\|_2,
\end{aligned}$$

for all $\alpha \in \mathcal{A}$.

This completes the proof.

$\square$

Note that the Jensen's equality claims that: If $\phi(V)$ is a convex function, then $\phi(E(V|X)) \leq E[\phi(V)|X]$. Here we choose $\phi(V) = V^2$ (which is a convex function) and $V = V(Z) = \frac{d\rho(Z,\alpha_0)}{d\alpha}(\alpha - \alpha_0)$.

Chapter 17

Endogeneity in Nonparametric Regression Models: Solutions

Exercise 17.1. Given that $E(u_i|Z_i) = 0$ we obtain from (17.21) that

$$E(Y_i|Z_i) = E[g(Z_i)|Z_i]. \tag{17.1}$$

Below we start from the left-hand-side of (17.25) and show that it equals the right-hand-side of (17.25). Using (17.1) we have

$$\begin{aligned}
&E[E(Y|Z)f_{xz}(w,,Z)] = E[E(g(X)|Z)f_{xz}(w,Z)] \\
&= E\left\{\left[\int g(x)f_{xz}(x,Z)f^{-1}(Z)dx\right] f_{xz}(w,Z)\right\} \\
&= \int f_z(z)[\int g(x)f_{xz}(x,z)f_z^{-1}(z)dx]f_{xz}(w,z)dz \\
&= \int \left[\int f_{xz}(x,z)f_{xz}(w,z)dz\right] g(x)dx \\
&= \int t(x,w)g(x)dx = (Tg)(w).
\end{aligned}$$

Exercise 17.2.

Chapter 18

Weakly Dependent Data: Solutions

Exercise 18.1. We use the short hand notation $a_{t,t+j} = Cov(K_{h,tx}, K_{h,t+j,x})$. By stationarity we know that $a_{t,t+s} = a_{t+j,t+j+s}$. Hence, we have

$$\begin{aligned}
\sum_{t=1}^{n-1}\sum_{j=1}^{n-t} a_{t,t+j} &= \sum_{t=1}^{n-1}[a_{t,t+1} + a_{t,t+2} + \cdots + a_{t,n}] \\
&= [a_{1,2} + a_{1,3} + \cdots + a_{1,n}] + [a_{2,3} + \cdots + a_{2,n}] + \cdots + [a_{n-1,n}] \\
&= (n-1)a_{1,2} + (n-2)a_{1,3} + \cdots + 2a_{1,n-1} + a_{1,n} \\
&= \sum_{j=1}^{n-1}(n-j)a_{1,1+j} = n\sum_{j=1}^{n-1}(1-j/n)a_{1,1+j}.
\end{aligned}$$

Exercise 18.2.

Exercise 18.3. Since the bias calculation is the same as in the independent data case, we only compute the variance term here. Letting $H_q = h_1 \ldots h_q$, we have

$$\begin{aligned}
Var(\hat{f}(x)) &= \frac{1}{n^2}\left\{\sum_{t=1}^{n} Var(K_{h,tx}) + 2\sum_{s>1}^{n}\sum_{t=1}^{n-1} Cov(K_{h,tx}, K_{h,sx})\right\} \\
&= \frac{1}{n^2}\left\{\sum_{t=1}^{n} Var(K_{h,tx}) + 2n\sum_{t=1}^{n-1}(1-t/n)Cov(K_{h,1x}, K_{h,1+tx})\right\} \\
&\leq \frac{1}{n^2}\left\{\sum_{t=1}^{n} Var(K_{h,tx}) + 8nM_n^{1/(1+\delta)}\sum_{t=1}^{n-1}(1-t/n)\beta^{\delta/(1+\delta)}(t)\right\} \\
&\leq \frac{1}{n}\{O(H_q^{-1}) + o(H_q^{-1})\} \\
&= O\left((nH_q)^{-1}\right)
\end{aligned}$$

because $M_n^{1/(1+\delta)} = O(H_q^{-2\delta/(1+\delta)}) = o(H_q^{-1})$ by taking $0 < \delta < 1$ as shown in the hint of this problem.

Hence,

$$MSE[\hat{f}(x)] = O\left(|h|^4 + (nh_1 \ldots h_q)^{-1}\right)$$

which implies that $\hat{f}(x) - f(x) = O_p\left(|h|^2 + (nH_q)^{-1/2}\right)$.

Chapter 19

Panel Data Models: Solutions

Exercise 19.1.

(i) T is finite and $N \to \infty$.

Using $\hat{f}(z) - f(z) = o_p(1)$, where $\hat{f}(z) = (NT)^{-1} \sum_{j=1}^{N} \sum_{s=1}^{T} K_{h,it,z}$, it is easy to see that that the leading term of $[\hat{g}(z) - g(z)] = [\hat{g}(z) - g(z)]\hat{f}(z)/\hat{f}(z)$ is $[\hat{g}(z) - g(z)]\hat{f}(z)/f(z) \equiv M(z)/f(z)$, where

$$
\begin{aligned}
M(z) &= [\hat{g}(z) - g(z)]\hat{f}(z) && (19.1) \\
&= (NT)^{-1} \sum_{j=1}^{N} \sum_{s=1}^{T} [g_{js} - g(z)] K_{h,js,z} \\
&\quad + (NT)^{-1} \sum_{j=1}^{N} \sum_{s=1}^{T} u_{js} K_{h,js,z} \\
&= M_1(z) + M_2(z). && (19.2)
\end{aligned}
$$

Obviously $Var[M_j(z)/f(z)] = Var[M_j(z)]/f(z)^2$, $j = 1, 2$. Below we will only compute $Var[M_j(z)]$.

Letting $A_i(z) = \sum_{t=1}^{T} [g_{it} - g(z)] K_{h,it,z}$, then $M_1(z) = (NT)^{-1} \sum_i A_i(z)$, we have

$$
\begin{aligned}
Var(A_i(z)) &= \sum_{t=1}^{T} Var[(g_{it} - g(z))K_{h,it,z}] + \sum_{t=1}^{T} \sum_{s \neq t}^{T} Cov\left[(g_{it} - g(z))K_{h,it,z}, (g_{is} - g(z))K_{h,is,z}\right] \\
&= O\left(\frac{|h|^2}{h_1 \ldots h_q}\right) + O(|h|^2)
\end{aligned}
$$

since T is finite ($T = O(1)$).

$Cov[A_i, A_j] = 0$ for $i \neq j$ by independence across the i-index. Hence,

$$
\begin{aligned}
Var[M_1(z)] &= \frac{1}{N^2 T^2} \sum_{i=1}^{N} \left\{ Var(A_i(z)) + \sum_{i=1}^{N} \sum_{j \neq i}^{N} Cov(A_i(z), A_j(z)) \right\} \\
&= O\left(N^{-2} N |h|^2 H_q^{-1}\right) = O\left((NH_q)^{-1} |h|^2\right) = o\left((NH_q)^{-1}\right).
\end{aligned}
$$

Letting $B_i(z) = \sum_{t=1}^{T} u_{is}K_{h,it,z}$, then $M_2(z) = (NT)^{-1}\sum_i B_i(z)$, we have (since T is finite)

$$\begin{aligned} Var(B_i(z)) &= \sum_{t=1}^{T} Var(u_{it}K_{h,it,z}) + \sum_{t=1}^{T}\sum_{s\neq t}^{T} Cov\left[u_{it}K_{h,it,z}, u_{is}K_{h,is,z}\right] \\ &= T\sigma^2(z)f(z)\kappa^q(h_1\dots h_q)^{-1} + o\left((h_1\dots h_q)^{-1}\right), \end{aligned}$$

where $\sigma^2(z) = E\left[u_{it}^2|z_{it} = z\right]$, $\kappa^q = \int K(v)^2dv$

For $j \neq i$, B_i and B_j are independent. Hence, we have

$$\begin{aligned} Var[M_2(z)] &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\left\{Var(B_i(z)) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} Cov(B_i(z), B_j(z))\right\} \\ &= \frac{\sigma^2(z)f(z)\kappa^q}{NTh_1\dots h_q} + o\left((nh_1\dots h_q)^{-1}\right). \end{aligned}$$

Finally, it is easy to show that $Cov[M_1(z).M_2(z)] = o\left((Nh_1\dots h_q)^{-1}\right)$. Hence, summarizing the above we have shown that

$$Var[M(z)/f(z)] = \frac{Var(M(z))}{f(z)^2} = \frac{\sigma^2(z)\kappa^q}{f(z)NTh_1\dots h_q} + o\left((nh_1\dots h_q)^{-1}\right).$$

Thus, we have shown that the asymptotic variance of $\hat{g}(z)$ is indeed the same as given in (19.4) of Li and Racine (2007). The asymptotic bias calculation is the same as i.i.d data case. By Liapunov's Central Limit Theorem, (19.4) follows.

Below we will first prove case (iii) because (ii) is a special case of (iii).

(ii) (Case (iii)) Both N and and T go to ∞.

In this case we assume that the data is ρ-mixing across the t-index (independent across the i-index). By the same derivation as in (i) and noting that

$$\begin{aligned} |Cov[(g_{it} - g(z))K_{h,it,z}, (g_{is} - g(z))K_{h,is,z}]| &\leq \rho(t-s)\sqrt{Var[(g_{it} - g(z))K_{h,it,z}]Var[(g_{is} - g(z))K_{h,is,z}} \\ &= \rho(t-s)Var[(g_{it} - g(z))K_{h,it,z}] \end{aligned}$$

by the ρ-mixing and stationary assumptions.

Hence, for $A_i(z) = \sum_{t=1}^{T}(g_{it} - g(z))K_{h,it,z}$, we have

$$\begin{aligned} Var(A_i(z)) &= \sum_{t=1}^{T} Var[(g_{it} - g(z))K_{h,it,z}] + \sum_{t}\sum_{s\neq t} Cov[(g_{it} - g(z))K_{h,it,z}, (g_{is} - g(z))K_{h,is,z}] \\ &= O\left(T|h|^2H_q^{-1}\right) + O\left(T|h|^2H_q^{-1}\right) = O\left(T|h|^2H_q^{-1}\right) \end{aligned}$$

by following the same proof for the ρ-mixing time series data case (since $\sum_{j=1}^{\infty}\rho(j)$ is finite).

Hence, we have

$$\begin{aligned} Var[M_1(z)] &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\{Var(A_i(z))+0\} \\ &= O\left(N^{-1}T^{-2}T|h|^2H_q^{-1}\right) \\ &= O\left(|h|^2(NTH_q^{-1})\right) \\ &= o\left((NTH_q)^{-1}\right). \end{aligned}$$

For $B_i = \sum_{t=1}^{T} u_{is}K_{h,it,z}$ and $M_2(z) = (NT)^{-1}\sum_i B_i(z)$, we have

$$\begin{aligned} Var(B_i(z)) &= \sum_{t=1}^{T} Var(u_{it}K_{h,it,z}) + \sum_{t=1}^{T}\sum_{s\neq t}^{T} Cov[u_{it}K_{h,it,z}, u_{is}K_{h,is,z}] \\ &= T\sigma^2(z)f(z)\kappa^q(h_1\dots h_q)^{-1} + o\left(T(h_1\dots h_q)^{-1}\right), \end{aligned}$$

by the same proof as the proof of Theorem 18.2 (see Li and Racine (2007, section 18.10.2, page 569)) because $\{z_{it}, u_{it}\}_{t=1}^{T}$ is a ρ-mixing time series data.

For $j \neq i$, B_i and B_j are independent. Hence, we have

$$\begin{aligned} Var[M_2(z)] &= \frac{1}{N^2T^2}\left\{\sum_{i=1}^{N} Var(B_i(z)) + 0\right\} \\ &= \frac{\sigma^2(z)f(z)\kappa^q}{NTh_1\dots h_q} + o\left((NTh_1\dots h_q)^{-1}\right). \end{aligned}$$

Finally, it is easy to show that $Cov[M_1(z), M_2(z)] = o\left((NTh_1\dots h_q)^{-1}\right)$. Hence, summarizing the above we have shown that

$$Var[M(z)/f(z)] = \frac{Var(M(z))}{f(z)^2} = \frac{\sigma^2(z)\kappa^q}{f(z)NTh_1\dots h_q} + o\left((NTh_1\dots h_q)^{-1}\right).$$

(iii) (Case (ii)) N is finite and $T\to\infty$. The derivation is exactly the same as in (iii) above. When N is finite we have $O\left((NTH_q^{-1})\right) = O\left((TH_q)^{-1}\right)$. Hence,

$$Var[M_1(z)] = O(|h|^2(TH_q)^{-1}) = o\left((TH_q)^{-1}\right), \qquad Cov[M_1(z), M_2(z)] = o\left((TH_q)^{-1}\right),$$

and

$$Var[M_2(z)] = \frac{\sigma^2(z)f(z)\kappa^q}{NTh_1\dots h_q} + o\left((NTh_1\dots h_q)^{-1}\right).$$

The remaining steps are the same as in the proof of (ii) above.

Exercise 19.2.

Exercise 19.3. Let $H_{i,[l-1]}$ be defined as in Li and Racine (2007, page 588). Then (19.40) can be written as

$$\begin{aligned} 0 &= \sum_{i=1}^{N} K_h(Z_{i1}, z) G_{i1} \left\{ -e'_{T-1}\Sigma^{-1} H_{i,[l-1]} + e'_{T-1}\Sigma^{-1} e_{T-1} \left[\hat{g}_{[l-1]}(Z_{i1}) - G'_{i1} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \right] \right\} \\ &\quad + \sum_{i=1}^{N} \sum_{t=2}^{T} K_h(Z_{it}, z) G_{it} \left\{ -c'_{T-1}\Sigma^{-1} H_{i,[l-1]} + c'_{T-1}\Sigma^{-1} c_{T-1} \left[\hat{g}_{[l-1]}(Z_{it}) - G'_{it} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \right] \right\} \\ &= -D_1 \begin{pmatrix} \alpha_0 \\ \alpha \end{pmatrix} + D_2 + D_3, \end{aligned}$$

where D_1, D_2 and D_3 are defined in (19.41) of Li and Racine (2007). Solving for $\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$ gives $\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} \hat{g}_{[l]}(z) \\ \hat{g}^{(1)}_{[l]}(z) \end{pmatrix} = D^{-1}(D_2 + D_3)$. Hence, (19.42) indeed gives the next step estimates of $\begin{pmatrix} g(z) \\ g^{(1)}(z) \end{pmatrix}$.

Exercise 19.4.

Exercise 19.5. From (19.63), (16.64) and noting that $\sqrt{N} w'u/n = \frac{1}{\sqrt{NT}} \sum_i W'_i u_i \xrightarrow{d} N(0, B)$ by the Lindeberg Central Limit Theorem and (16.64), we have

$$\begin{aligned} \sqrt{N}(\tilde{\alpha} - \alpha) &= \sqrt{N}(v'ww'v)^{-1} v'ww'u[(v'w/n)(w'v)/n]^{-1}(v'w/n)\sqrt{N} w'u/n \\ &\xrightarrow{d} (A'A)^{-1} A' N(0, B) = Q^{-1} A' N(0, B) = N(0, V), \end{aligned}$$

where $V = Q^{-1} A' B A Q^{-1}$.

Exercise 19.6.

Exercise 19.7. From Li and Racine (2007, page 622) we know that $\Omega(z)[\hat{g}_{[l]}(z) - g(z)] = A_N + o_p(\eta_N)$, $\eta_N = \sum_{r=1}^{q} h_r^2 + (N h_1 \dots h_q)^{-1/2})$, $A_N = A_{1N} + A_{2N}$, $A_{1N} = A_{1N1} + A_{1N2}$ with $E(A_{1N1}) = (\kappa_2/2)\Omega(z) \sum_{r=1}^{q} h_r^2 g_{rr}(z) + o_p(\sum_{r=1}^{q} h_r^2)$ and $E(A_{1N2}) = 0$ (since $E(u_{is}|Z) = 0$).

Hence, we only need to consider A_{2N} which is defined by (19.102):

$$A_{2N} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} K_h(Z_{it}, z) \sum_{s=1}^{T} \sigma^{ts} [\hat{g}(Z_{is}) - g(Z_{is})],$$

where $\hat{g}(Z_{is})$ is the first step estimator of $g(Z_{is})$ that ignores the variance structure Σ. This is a standard local linear estimator for $g(Z_{is})$. Hence, from the result of Chapter 2 we know that the bias is given by

$$E\left[\hat{g}(Z_{is}) - g(Z_{is}) | Z_{it} = z\right] = \frac{\kappa_2}{2} \sum_{r=1}^{q} E\left[g_{rr}(Z_{is}) | Z_{it} = z\right] + o_p\left(\sum_{r=1}^{q} h_r^2\right).$$

Substituting the above result into A_{2N} and combining it with (19.101) we obtain the leading bias for $\Omega(z)[\hat{g}_{[1]}(z) - g(z)] \equiv A_N + (s.o.) = A_{1N} + A_{2N} + (s.o.)$ given by

$$\begin{aligned}
&\frac{\kappa_2}{2}\Omega(z)\sum_{r=1}^{q} h_r^2 g_{rr}(z) + \frac{\kappa_2}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{q} h_r^2 E[g_{rr}(Z_{is})|Z_{it} = z] \\
&= \Omega(z)\left\{\frac{\kappa_2}{2}\sum_{r=1}^{q} h_r^2\left[g_{rr}(z) + \Omega(z)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} E[g_{rr}(Z_{is})|Z_{it} = z]\right]\right\} + (s.o.).
\end{aligned}$$

The above is the leading bias of $\Omega(z)[\hat{g}_{[1]}(z) - g(z)]$. Premultiplying the above by $\Omega(z)^{-1}$ gives the leading bias for $\hat{g}_{[1]}(z) - g(z)$ as given in (19.104).

Chapter 20

Topics in Applied Nonparametric Estimation: Solutions

Exercise 20.1. This follows from

$$\int_{t\Delta}^{(t+1)\Delta}(u-t\Delta)du = \frac{1}{2}u^2|_{u=t\Delta}^{(t+1)\Delta} - t\Delta^2 = (1/2)\left[2t\Delta^2+\Delta^2\right] - t\Delta^2 = \Delta^2/2.$$

Exercise 20.2.

Bibliography

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* **14**, 1–16.

Härdle, W., P. Hall and H. Ichimura. (1993). Optimal smoothing in single index models. *Annals of Statistics* **21**, 157–178.

Hayfield, Tristen and Jeffrey S. Racine. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software* **27**(5).
URL: *http://www.jstatsoft.org/v27/i05/*

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71–120.

Li, Q. (1996). On the root-n-consistent semiparametric estimation of partially linear models. *Economics Letters* **51**, 277–285.

Li, Q. and J. Racine. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Liu, Z. and X. Lu. (1997). Root-n-consistent semiparametric estimation of partially linear models based on k-nn method. *Econometric Reviews* **16**, 411–420.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* **79**, 147–168.

Pagan, A. and A. Ullah. (1999). *Nonparametric econometrics*. New York: Cambridge University Press.

Wang, J. and L. Yang. (2005). Efficient and fast spline-backfitted kernel smoothing of additive regression model. Statistics and probability RM 638 Michigan State University.