Semiparametic Regression - Assignment 5

Szymon Armata | 341593 | Data Science

15 November 2022

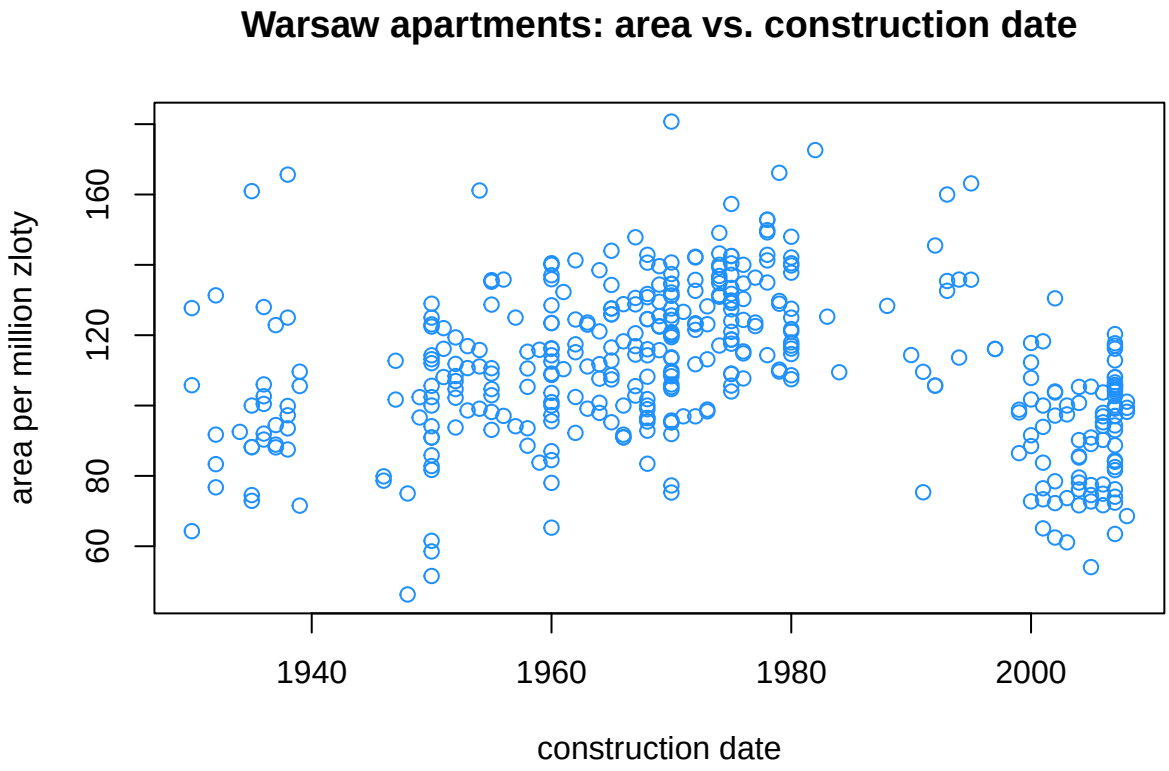
Contents

1 Plot of the data

```
library(nlme)
library(HRW)
data(WarsawApts)

data(WarsawApts)
x <- WarsawApts$construction.date
y <- WarsawApts$areaPerMzloty
```

```
plot(x,y,
     xlab = "construction date",
     ylab = "area per million zloty", col="dodgerblue",
     main = "Warsaw apartments: area vs. construction date")
```



2 Task 3.

2.1 a

```
area.perMz <- WarsawApts$areaPerMzloty
const.date <- WarsawApts$construction.date
```

2.2 b

```
numObs <- length(const.date)
X <- cbind(rep(1,numObs), const.date)
```

2.3 c

```
numIntKnots <- 35
intKnots <- quantile(unique(const.date),
seq(0, 1,
length = numIntKnots + 2))[-c(1,numIntKnots + 2)]
Z <- outer(const.date, intKnots, "-")
Z <- Z * (Z > 0)
```

2.4 d

```
dummyId <- factor(rep(1, numObs))
Z.sm <- list(dummyId = pdIdent(~ -1 + Z))
fit <- lme(area.perMz ~ -1 + X, random = Z.sm)
```

3 Task 4.

3.1 a

```
ng <- 1001
range.date <- range(const.date)
dategrid <- seq(range.date[1], range.date[2], length = ng)
Xg <- cbind(rep(1, ng), dategrid)
Zg <- outer(dategrid, intKnots, "-")
Zg <- Zg * (Zg > 0)
```

3.2 b

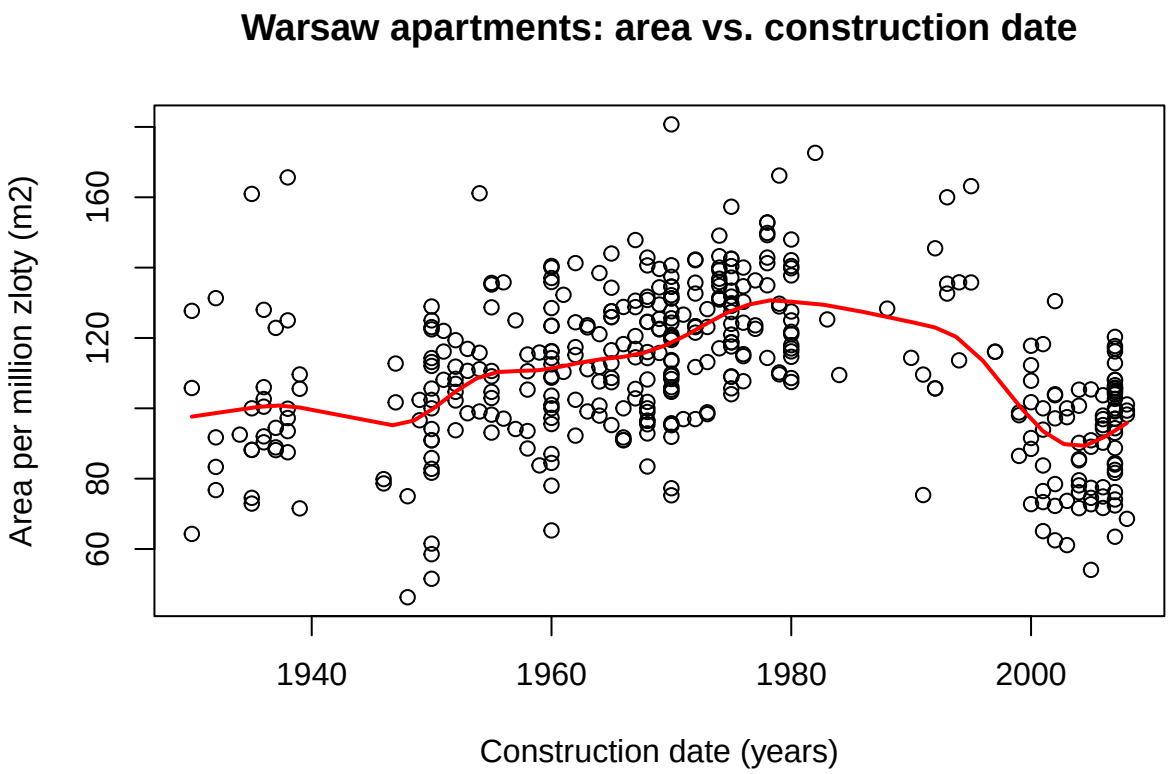
```
betaHat <- as.vector(fit$coef$fixed)
uHat <- as.vector(fit$coef$random[[1]])
```

3.3 c

```
fhat <- Xg %*% betaHat + Zg %*% uHat
```

3.4 d

```
plot(const.date, area.perMz,
xlab = "Construction date (years)",
ylab = "Area per million zloty (m2)",
main = "Warsaw apartments: area vs. construction date")
lines(dategrid, fhat, lwd = 2, col = "red")
```



Even though there are times when there is a shortage of data (for example, the years 1940-1945 or 1980–1995) and the line appears to be overfitted, it can be seen that the fitting line represents the trend in apartment prices.