# Semiparametric regression - Homework 5

Klaudia Weigel

The data used in this exercise is the `WarsawApts` data. We will fit a semiparametric model relating the "area per million zloty" to the "construction date". We will define the semiparametric model with a linear mixed model:

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} u_k z_k(x_i) + \epsilon_i,$$

where $y$ is the value of "area per million zloty" and $x$ is the "construction date", $u_k$ are independent random variables with $u_k \sim N(0, \sigma_u^2)$. As $z_k$ we take a function $z_k(x) = [(x - \kappa_k)_+]^2$.

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

numObs <- length(const.date)
# Design matrix
X <- cbind(rep(1,numObs), const.date)
# Creating the random design matrix "Z" via basis function definition with
# a user-specified number of knots and evaluated at the predictor const.date
# values
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)

# Setting up the linear mixed model defining the semiparametric model
dummyId <- factor(rep(1, numObs))
Z.sm <- list(dummyId = pdIdent(~ -1 + Z))
fit <- lme(area.perMz ~ -1 + X, random = Z.sm)
```

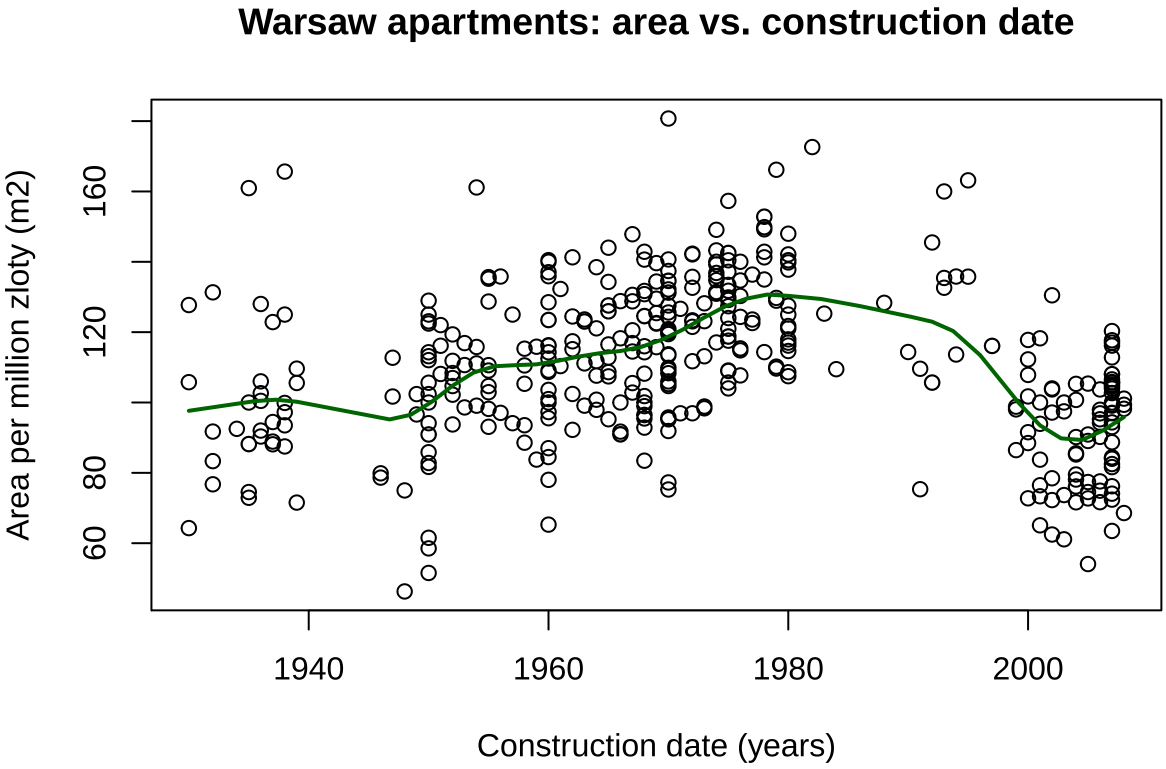
```
# (a) Setting up the grid values together with the fixed and
# random design matrices
ng <- 1001
range.date <- range(const.date)
dategrid <- seq(range.date[1], range.date[2], length = ng)
Xg <- cbind(rep(1, ng), dategrid)
Zg <- outer(dategrid, intKnots, "-")
Zg <- Zg * (Zg > 0)

# (b) Extracting the model parameters
betaHat <- as.vector(fit$coef$fixed)
uHat <- as.vector(fit$coef$random[[1]])

# (c) Estimated semiparametric model fit
```

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

# (d) Plot of the fitted curve
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 = "darkgreen")
```



We see that the fitted line is good represantation of the trends in the dataset.