Delivery 4: AMA

Marc Falcón Barau, Julian Fransen,Victor Garcia Pizarro

2024-10-03

Table of Contents

Import Data

We are using the Aircraft data from the R package sm. This dataset records six characteristics of aircraft designs that emerged during the twentieth century.

```
library(sm)

## Warning: package 'sm' was built under R version 4.2.3

## Package 'sm', version 2.2-6.0: type help(sm) for summary information

library(KernSmooth)

## KernSmooth 2.23 loaded
## Copyright M. P. Wand 1997-2009

data(aircraft)
# help(aircraft)
attach(aircraft)

lgPower <- log(Power)
lgSpan <- log(Span)
lgLength <- log(Length)
lgWeight <- log(Weight)
lgSpeed <- log(Speed)
lgRange <- log(Range)
```

Now, we are interested in creating a heteroscedastic regression model with $x = \text{Yr}$ and $y = \text{lgWeight}$. To do this, we need to calculate the error term ϵ, which must be approximated from the variance function $\sigma^2(x)$, as it is not constant in this case. To find this function, we use the following approach:

1. Calculate the expected mean function $m(x_i)$.
2. Transform the estimated residuals $(z = \log(\epsilon_i^2))$.
3. Estimate the function $q(x)$.
4. Calculate the variance function by computing $\exp\big(q(x)\big)$.

First, we will use the `locpolreg()` function to create nonparametric models and apply cross-validation (CV) to select the optimal bandwidth. Then, we will use `sm.regression()` and `dpill()` to estimate the bandwidth.

Approach 1: locpolreg() and CV approach

1. Fit a nonparametric regression to data (xi, yi) and save the estimated values m(xi).

```
source("locpolreg.R")

optimal_h1a <- h.select(x=Yr, y=lgWeight, method="cv")
model1a <- locpolreg(x = Yr, y = lgWeight, h = optimal_h1a,
doing.plot=FALSE)

# Get the estimated regression values
estimated_m <- model1a$mtgr
```

2. Transform the estimated residuals ϵi = yi − m(xi):

```
residuals <- lgWeight - estimated_m
stimated_residuals <- log(residuals^2)
```

3. Fit a nonparametric regression to data (xi, zi) and call the estimated functionˆq(x). Observe that ˆq(x) is an estimate of log σ2(x).

```
optimal_h2a <- h.select(x = Yr, y = stimated_residuals, method="cv")
model2a <- locpolreg(x = Yr, y = stimated_residuals, h = optimal_h2a,
doing.plot=FALSE)
```

4. Estimate σ2(x)

```
estimated_variance <- exp(model2a$mtgr)
```

5. Final plot

Finally, we combine all the elements that we have estimated: residuals, mean, and variance.

```
# Calculate and plot ϵ²ᵢ against xᵢ
plot(Yr, residuals ^ 2,
     xlab = "Yr",
     ylab = "lgWeight",
     main = "locpolreg() and CV",
     col = "blue",
     pch = 16)  # Set the same y-axis range for both

# Add a legend with all elements
legend("topleft",
       legend = c("ϵ²", "σ²(x)",
                  "m(x)", "95% CI"),
```

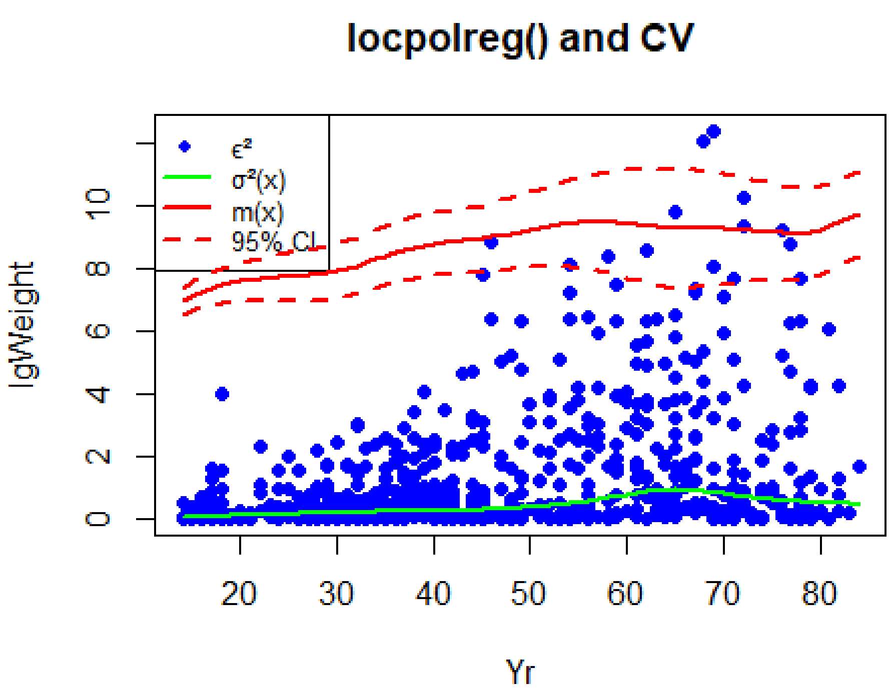
```r
       col = c("blue", "green", "red", "red"),
       pch = c(16, NA, NA, NA),  # points are residuals
       lty = c(NA, 1, 1, 2),      # Line type for other elements
       lwd = c(NA, 2, 2, 2),      # Line width
       cex = 0.8,
       xpd = TRUE
)

# Superpose estimated σ²(x)
lines(Yr, estimated_variance, col = "green", lwd = 2)

# Superpose estimated mean with 95% CI
lines(Yr, estimated_m, col = "red", lwd = 2)
estimated_sd = sqrt(estimated_variance)

lower_band <- estimated_m - 1.96 * estimated_sd
upper_band <- estimated_m + 1.96 * estimated_sd

lines(Yr, lower_band, col = "red", lty = 2, lwd = 2)
lines(Yr, upper_band, col = "red", lty = 2, lwd = 2)
```



Approach 2: sm.regresion and dpill() approach

1. Fit a nonparametric regression to data (xi, yi) and save the estimated values m(xi).

This approach creates a different number of estimated points than Yr, so it is necessary to perform linear interpolation to maintain the same number of points and ensure they can be plotted later.

```
optimal_h1b <- dpill(x = Yr, y = lgWeight)
model1b <- sm.regression(x = Yr, y = lgWeight, h = optimal_h1b,
display="none")

# Get the estimated regression values
estimated_m <- model1b$estimate
estimated_m_interpolated <- approx(model1b$eval.points, model1b$estimate,
xout = Yr)$y
```

2. Transform the estimated residuals ϵi = yi – m(xi):

```
residuals <- lgWeight - estimated_m_interpolated
stimated_residuals <- log(residuals^2)
```

3. Fit a nonparametric regression to data (xi, zi) and call the estimated functionˆq(x). Observe that ˆq(x) is an estimate of log σ2(x).

```
optimal_h2b <- dpill(x = Yr, y = stimated_residuals)
model2b <- sm.regression(x = Yr, y = stimated_residuals, h = optimal_h2b,
display="none")

## missing data are removed

estimated_q_interpolated <- approx(model2b$eval.points, model2b$estimate,
xout = Yr)$y
```

4. Estimate σ2(x)

```
estimated_variance <- exp(estimated_q_interpolated)
```

5. Final plot

Again, we plot all elements we have calculated

```
# Calculate and plot ϵ²ᵢ against xᵢ
plot(Yr, residuals ^ 2,
     xlab = "Yr",
     ylab = "",
     main = "sm.regression() and dpill()",
     col = "blue",
     pch = 16)  # Set the same y-axis range for both

legend("topleft",
```

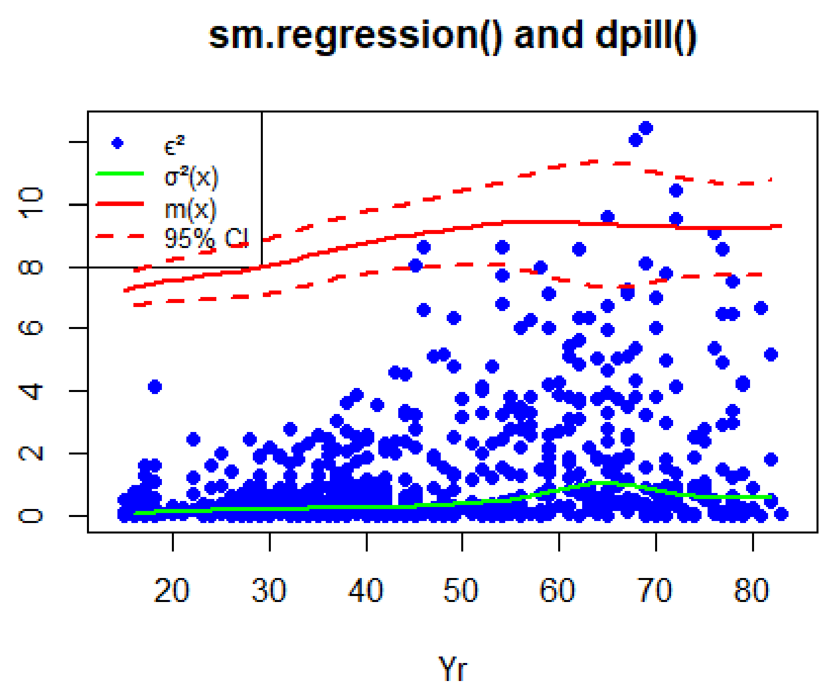
```r
       legend = c("ϵ²", "σ²(x)",
                  "m(x)", "95% CI"),
       col = c("blue", "green", "red", "red"),
       pch = c(16, NA, NA, NA),  # points are residuals
       lty = c(NA, 1, 1, 2),     # Line type for other elements
       lwd = c(NA, 2, 2, 2),     # Line width
       cex = 0.8,
       xpd = TRUE
)

# Superpose estimated σ²(x)
lines(Yr, estimated_variance, col = "green", lwd = 2)

# Superpose estimated mean with 95% CI
lines(Yr, estimated_m_interpolated, col = "red", lwd = 2)
estimated_sd = sqrt(estimated_variance)

lower_band <- estimated_m_interpolated - 1.96 * estimated_sd
upper_band <- estimated_m_interpolated + 1.96 * estimated_sd

lines(Yr, lower_band, col = "red", lty = 2, lwd = 2)
lines(Yr, upper_band, col = "red", lty = 2, lwd = 2)
```



Conclusions

Both approaches are very similar, exhibiting low variance close to zero, as expected. The residuals are also well-distributed around zero, particularly in the earlier years. Overall, we observe that variance increases over time, resulting in wider confidence intervals (CIs) and a bigger residuals.