STATS 305A : HW5

Problem 2

(a) We want to fit the model:

$$f(x, t) = \beta_0(t) + \beta_1(t)x + \varepsilon_t$$

where t is a value of the variable T, and x a value of the predictor X.

In order to estimate (β_0, β_1) with respect to T, we will use a polynomial splines with a truncated power basis.

By choosing $M-4$ knots evenly spread over the values of T we would have the basis:

$$1, x, x^2, x^3, (x-h_1)_+^3, \dots, (x-h_{M-4})_+^3$$

and also we could define the basis matrix of splines in T: H_T ($m \times M$ matrix)

s.t.:

$$\begin{cases} \underline{\beta_0} = H_T \underline{\theta_0} \\ \underline{\beta_1} = H_T \underline{\theta_1} \end{cases}$$

In other words, our model

writes:

$$\mathbb{E}(Y \mid T, X) = \sum_{m=1}^{M} \theta_{0m} h_m(t) + \sum_{j=1}^{p} X_j \sum_{m=1}^{M} \theta_{jm} h_m(t)$$

So, when $p = 1$ like in our case, there are exactly:

$$\boxed{2 \times M}$$

parameters to estimate which are:

$\theta_{01}, \dots, \theta_{0M}, \theta_{11}, \dots, \theta_{1M}$.

For $i \in [\![1, n]\!]$, as the model

writes:

$$y_i = \sum_{m=1}^{n} \theta_{0m} h_m(t_i)$$

$$+ \sum_{m=1}^{M} \theta_{1m} h_m(t_i)\, x_i$$

we can see it as a regression model with covariates:

$$h_1(T), \dots, h_M(T), h_1(T)X,$$

$$\dots, h_M(T)X$$

So by writing:

$$\tilde{X} = \left(h_1(T), \dots, h_M(T), Xh_1(T), \dots, Xh_M(T)\right)$$

we can fit the linear regression:

$$y = \hat{\tilde{X}} \alpha + \varepsilon$$

where α is a vector of size $2n$.

(d)

- The slope function writes:

$$\hat{\beta}_1(t) = H_{T=t} \times \hat{\theta}_1 \, , \forall t$$

where

$$\hat{\theta}_1 = (\hat{\theta}_{11}, \dots, \hat{\theta}_{n1})$$

defined as `theta.1` in

the written handout.

So we have:

$$\hat{\beta}_1(t) = \sum_{m=1}^{M_1} \hat{\theta}_{m1} h_m(t).$$

Besides, for $m > 2$, $h_m(t)$ is polynomial in t with degree at least 3. Consequently, in order to test whether $\hat{\beta}_1(\cdot)$ is linear in E, we need to compute a

F-test

with null hypothesis :

$H_0 : \beta_{31} = \beta_{41} = \dots = \beta_{M1} = 0$

The test statistic can be defined:

$$\boxed{F_{stat} = \frac{(RSS_{restricted} - RSS_{ur})/n+2}{RSS_{ur}/(n-2M)}}$$

where $RSS_{restricted}$ are the residuals sum of squares for the restricted model with only $n+2$ predictors.

• The function form of this model is:

$$\mathbb{E}(Y|X,T) = \beta_0(T) + \overbrace{h_1(t)}^{=1}\theta_{11}\, x_1 + \underbrace{h_2(t)}_{=t}\theta_{21}\, x_1$$

$$= \beta_0(T) + (\theta_{11}(t) + \theta_{21} \cdot t)\, x_1$$

The answers are right after the code part.