## 9.2 Correct regression function

7

8

**Assumed model** $M: Y/Z \sim (\underline{X}\beta, \sigma^2 I_n)$

$\varepsilon = Y - X\beta$ satisfy $\mathbb{E}(\varepsilon/Z) = 0$

$var(\varepsilon/Z) = \sigma^2 I$

---

(A1) correct regression function

$\mathbb{E}(Y/Z) \in \mathcal{M}(\underline{X})$ $\equiv \mathbb{E}(Y/Z) = X\beta$

was chosen by us for some $\beta \in \mathbb{R}^k$

$\Leftrightarrow$ $\mathbb{E}(\varepsilon/Z) = 0_n$ $(\Rightarrow \mathbb{E}(\varepsilon) = 0_n)$

---

(A1) $\Rightarrow$ $\boxed{\mathbb{E}(U/Z) = 0_n}$

$U = Y - \hat{Y}$ residuals

This property checked on residual plots.

---

**Assumption** for the rest of this section:

$X = (\mathbb{1}, X^1, \dots, X^{k-1})$, rank$(X) = k < n$

$\hat{=} X^0$

(FULL-RANK)

**Will follow:** methods to determine which $X^j$ might be responsible for $\mathbb{E}(\varepsilon/Z) \neq 0_n$

## 9.2.1 Partial residuals

NOTATION: Model with a removed $j^{th}$ regressor

$\mathbb{X}^{(-j)} :=$ matrix $\mathbb{X}$ without the column $X^j$

$\beta^{(-j)} := (\beta_0, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_{k-1})^T$

---

$M^{(-j)}: \ Y|Z \sim (\mathbb{X}^{(-j)}\beta^{(-j)}, \sigma^2 I_n)$

$\mathbb{M}^{(-j)} = I_n - \mathbb{X}^{(-j)}(\mathbb{X}^{(-j)T}\mathbb{X}^{(-j)})^{-1}\mathbb{X}^{(-j)T}$ (residual projection matrix for model $M^{(-j)}$)

$U^{(-j)} = \mathbb{M}^{(-j)}Y$ (residuals of model $M^{(-j)}$)

---

We are assuming $\mathbb{X} = (\underline{\mathbb{1}}, X^1, \dots, X^{k-1})$, $\text{rank}(\mathbb{X}) = k$

$\Rightarrow \text{rank}(\mathbb{X}^{(-j)}) = k-1$

$\Rightarrow$ (i) $X^j \notin \mathcal{M}(\mathbb{X}^{(-j)})$

(ii) $X^j \neq 0_n$

(iii) $X^j$ is not a multiple of a vector $\mathbb{1}_n$.

---

Model $M: Y|Z \sim (\mathbb{X}\beta, \sigma^2 I_n)$ is <u>model with added regressor</u> as compared to model $M^{(-j)}: Y|Z \sim (\mathbb{X}^{(-j)}\beta^{(-j)}, \sigma^2 I_n)$.

---

Remember <u>Lemma 9.1</u> (in its notation)

$c_g = (V^T \mathbb{M} V)^{-} V^T U$

$b_g = b - (X^T X)^{-} X^T V c_g$

NOW:

$V \equiv X^j$

$X \equiv \mathbb{X}^{(-j)}$

$\mathbb{M}$ = resid. project matrix of $M^{(-j)}$

$U$ = residuals of $M^{(-j)}$

$c_g = \hat{\beta}_j$ in $M$

$b_g = \hat{\beta}^{(-j)}$ in $M$

$b = \hat{\beta}^{(-j)}$ in $M^{(-j)}$

We will use Lemma 9.1 to express $\hat{\beta}_j$ using quantities from $M^{(-j)}$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (\equiv \text{LSE of } \beta \text{ in model } M)$$

$$= (\underbrace{\underbrace{\hat{\beta}_0, \dots, \hat{\beta}_{j-1}}, \underbrace{\hat{\beta}_j}_{= c_g}, \underbrace{\hat{\beta}_{j+1}, \dots, \hat{\beta}_{k-1}}}_{b_g})^T$$

invert ?

Lemma 9.1 $\Rightarrow$

$$c_g = \hat{\beta}_j = (X^{jT} M^{(-j)} X^j)^{-1} X^{jT} U^{(-j)}$$

- can we invert ?

$$X^{jT} M^{(-j)} X^j = X^{jT} M^{(-j)T} M^{(-j)} X^j =$$

$$= \| M^{(-j)} X^j \|^2$$

- $X^j \notin \mathcal{M}(X^{(-j)})$, $X^j \neq 0 \Rightarrow \underbrace{M^{(-j)} X^j}_{\text{projection of } X^j \text{ to } \mathcal{M}(X^{(-j)})^\perp} \neq 0_n$

$$\Rightarrow \| M^{(-j)} X^j \|^2 > 0$$

- moreover, $X^{jT} M^{(-j)} X^j$ is scalar and hence we can write

$$\boxed{\hat{\beta}_j = \frac{X_j^T U^{(-j)}}{X^{jT} M^{(-j)} X^j}}$$

## Def 9.1 Partial residuals

10

A vector of $j$th partial residuals of model $M$ is a vector

$$U^{part,j} = U + \hat{\beta}_j X^j = \begin{pmatrix} U_1 + \hat{\beta}_j X_{1,j} \\ \vdots \\ U_n + \hat{\beta}_j X_{n,j} \end{pmatrix}.$$

---

NOTE: We have

$$\begin{aligned} U^{part,j} &= U + \hat{\beta}_j X^j \\ &= Y - (X\hat{\beta} - \hat{\beta}_j X^j) \\ &= Y - (\underbrace{\hat{Y} - \hat{\beta}_j X^j}_{\text{"partial" fitted values}}) \end{aligned}$$

---

## Lemma 9.2 Property of partial residuals

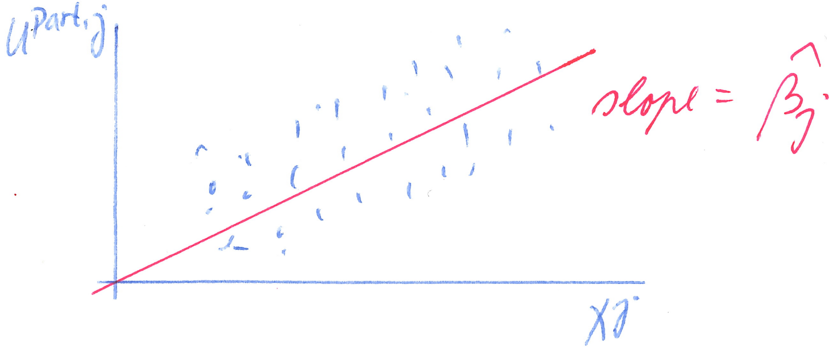
11

Let $Y|Z \sim (X\beta, \sigma^2 I_n)$, $\mathrm{rank}(X_{n\times k}) = k$, $X^0 = \mathbb{1}_n$, $\beta = (\beta_0, \dots, \beta_{k-1})^T$. Let $\hat{\beta}_j$ be the LSE of $\beta_j$, $j \in \{1, \dots, k-1\}$. Let us consider a linear model (regression line with covariates $X^j$) with

- the $j$th partial residuals $U^{part,j}$ as response;
- a matrix $(\mathbb{1}_n, X^j)$ as the model matrix;
- regression coefficients $\gamma_j = (\gamma_{j0}, \gamma_{j1})^T$.

The least squares estimators of parameters $\gamma_{j0}$ and $\gamma_{j1}$ are

$$\hat{\gamma}_{j0} = 0, \quad \hat{\gamma}_{j1} = \hat{\beta}_j.$$




Proof:

- $U^{part,j} = U + \hat{\beta}_j X^j$
- To get LSE of $\gamma_j = (\gamma_{j,0}, \gamma_{j,1})^T$, we must minimize the sum of squares

$$\| U^{part,j} - \gamma_{j,0}\mathbb{1} - \gamma_{j,1} X^j \|^2 =$$

$$= \| U - \{ \gamma_{j,0}\mathbb{1} + (\gamma_{j,1} - \hat{\beta}_j) X^j \} \|^2 = (*)$$

$U \in \mathcal{M}(X)^\perp$ $\quad$ $\mathbb{1} \in \mathcal{M}(X)$, $X^j \in \mathcal{M}(X)$

Pythagoras

$$(*) = \|U\|^2 + \| \gamma_{j,0}\mathbb{1} + (\gamma_{j,1} - \hat{\beta}_j) X^j \|^2$$

$$\geq \|U\|^2$$

equality here $\iff$ $\gamma_{j,0} = 0$, $\gamma_{j,1} = \hat{\beta}_j$

$X^j \neq 0$

$\mathbb{1}$ and $X^j$ linearly independent

Hence $\hat{\gamma}_{j,0} = 0$, $\hat{\gamma}_{j,1} = \hat{\beta}_j$ ∎

(Obvious) notation:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \overline{X^j} = \frac{1}{n}\sum_{i=1}^{n} X_{ij}, \quad \overline{U}^{part,j} = \frac{1}{n}\sum_{i=1}^{n} U_i^{part,j}$$

If $X^0 = \mathbb{1}_n$ (model with intercept) then

$$\underset{\substack{\uparrow \\ \text{Lemma 7.2}}}{0 =} \sum_{i=1}^{n} U_i = \sum_{i=1}^{n} \left(U_i^{part,j} - \hat{\beta}_j X_{ij}\right)$$

$$\frac{1}{n}\sum_{i=1}^{n} U_i^{part,j} = \hat{\beta}_j \left(\frac{1}{n}\sum_{i=1}^{n} X_{ij}\right)$$

$$\overline{U}^{part,j} = \hat{\beta}_j \overline{X^j}$$

Def 9.2 Shifted partial residuals

A vector of $j$th response-mean partial residuals of model $M$ is a vector

$$U^{part,j,Y} = U^{part,j} + \left(\overline{Y} - \hat{\beta}_j \overline{X^j}\right)\mathbb{1}_n.$$

A vector of $j$th zero-mean partial residuals of model $M$ is a vector

$$U^{part,j,0} = U^{part,j} - \hat{\beta}_j \overline{X^j}\mathbb{1}_n.$$

$\hookrightarrow$ residuals(·, type = "partial") in R

## Interpretation of partial residuals

13

→ see slide

## Use of partial residuals

→ see slide

---

## Example: Cars2004nh

14

$$\text{consumption} \sim \underbrace{\log(\text{weight}) + \text{engine size} + \text{horsepower}}_{\text{assumed model}}$$

---

15

Useful to realize difference between marginal and partial association

$Z$: covariate 1

$V$: other covariates

Let $\mathbb{E}(Y|Z=z, V=v) = \beta_0 + \beta_1 z + (\beta^V)^T v$

$$\Rightarrow \mathbb{E}(Y|Z=z) = ?$$

$$\mathbb{E}(Y|Z=z) = \int \mathbb{E}(Y|Z=z, V=v) f(v|z)\, dv =$$

assume that $V|Z=z$ is continuous with a density $f(v|z)$

$$= \int (\beta_0 + \beta_1 z + (\beta^V)^T v) f(v|z)\, dv =$$

$$= \beta_0 + \beta_1 z + (\beta^V)^T \underbrace{\mathbb{E}(V|Z=z)}_{\text{some function of } z\,!}$$

---

That is, even if $\mathbb{E}(Y|Z=z, V=v)$ is linear in $z$,
(multiple regression)

$\mathbb{E}(Y|Z=z)$ is not necessarily linear in $z$.

Even if $\mathbb{E}(Y|Z=z)$ appears to be linear in $z$, the respective slope may differ from the slope from $\mathbb{E}(Y|Z=z, V=v)$.

SUMMARY: 15

- To build a multiple regression model, separate simple regressions are pretty useless (they do not provide much information on how the covariate Z should be parameterized in the multiple regression model).
- Useful parameterization for the covariate Z can be proposed from the partial residual plot (provided that the remaining covariates are already correctly included in the model ... )

→ building a good model usually needs some iterations and it is time consuming...

it shows "clean" association between Y and Z (being polished from influence of V - other covar.)

Conclusion for Cars2004nh: 16-17

- all three covariates seem to be pretty correctly parameterized for purpose of modelling $E(Y|Z_1, Z_2, Z_3)$
- both engine size and horsepower would need some transformation in case a marginal association is of interest

Example: Policie 18-20

Simpson's paradox caused by mutual associations among fat, height, weight.

## 9.2.2 Test for linearity of the effect

AIM: Statistical test to evaluate whether the effect of $X^j$ is significantly non-linear (in a multiple regres. model).

ASSUME (with loss of generality): $X = (1, X^0, X^j) \equiv M$

---



## More general parameterization of the $j^{th}$ regressor

- consider a (larger) model $M_g : (1, X^0, \mathbb{V})$ such that $\text{rank}(\mathbb{V}) \geq 2$

$$\mathcal{M}(X^j) \subsetneq \mathcal{M}(\mathbb{V})$$

$\mathbb{V} \equiv$ more general parameterization of certain covariate than that provided by regressor $X^j$

$\Longrightarrow$ submodel testing

Submodel $M : (1, X^0, X^j) \equiv X$

(Larger) model $M_g : (1, X^0, \mathbb{V})$

---

F-test rejects the submodel $M$

$\Longrightarrow$ effect of $X^j$ is significantly NOT linear

---

Examples will follow for a model

$$M: \mathbb{E}(Y|\ldots) = \beta_0 + \beta_1 \log(w) + \beta_2 z + \beta_3 \underset{X^j}{x}$$

consumption, weight, engine size, horse power

## Possibilities on how to choose $\mathbb{V}$

$$X^j = \begin{pmatrix} X_{1j} \\ \vdots \\ X_{nj} \end{pmatrix} \text{ (horsepower)}$$

(a) $\mathbb{V} \equiv$ polynomial

e.g. $\mathbb{V} = \begin{pmatrix} X_{1j} & X_{1j}^2 \\ \vdots & \vdots \\ X_{nj} & X_{nj}^2 \end{pmatrix}$

R output $\boxed{23}$

basic resid. plots $\boxed{24}$

---

(b) $\mathbb{V} \equiv$ regression spline

$$\mathbb{V} = \begin{pmatrix} B_1(X_{1j}) & \cdots & B_k(X_{1j}) \\ \vdots & & \vdots \\ B_1(X_{nj}) & \cdots & B_k(X_{nj}) \end{pmatrix}$$

$B_1, \ldots, B_k$: chosen spline basis that could possibly parameterize $X^j$

---

- choice of knots for spline basis $\boxed{25}$
- estimated spline coefficients → not of interest

basic resid. plots $\boxed{26}$

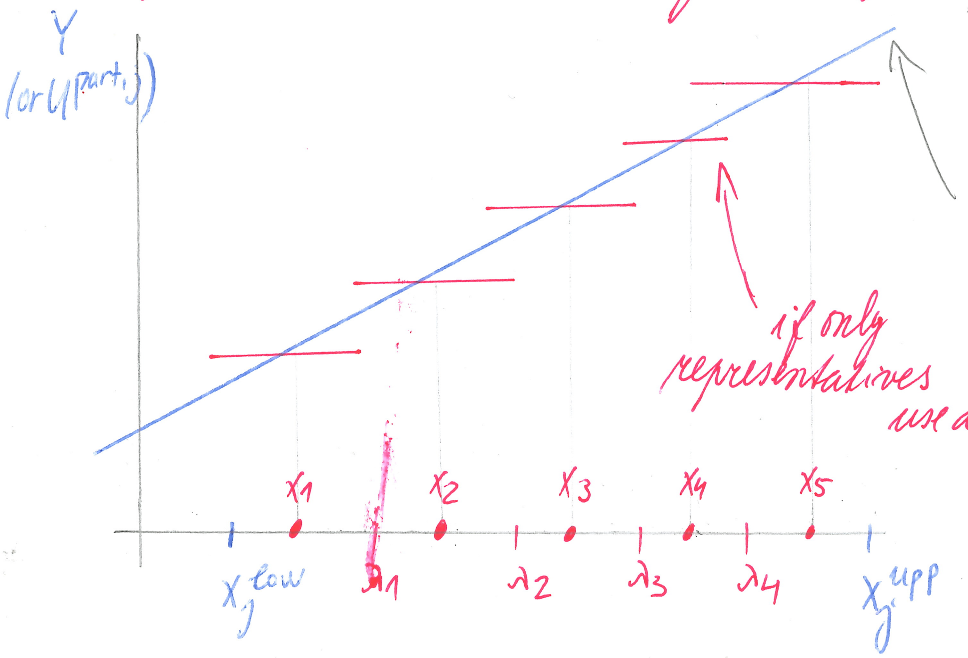
- F-test $\boxed{27}$

- Fitted spline

$$\equiv \hat{\beta}_1 B_1(x) + \ldots + \hat{\beta}_6 B_6(x)$$

– usage?

# Categorization of the $j^{th}$ regressor

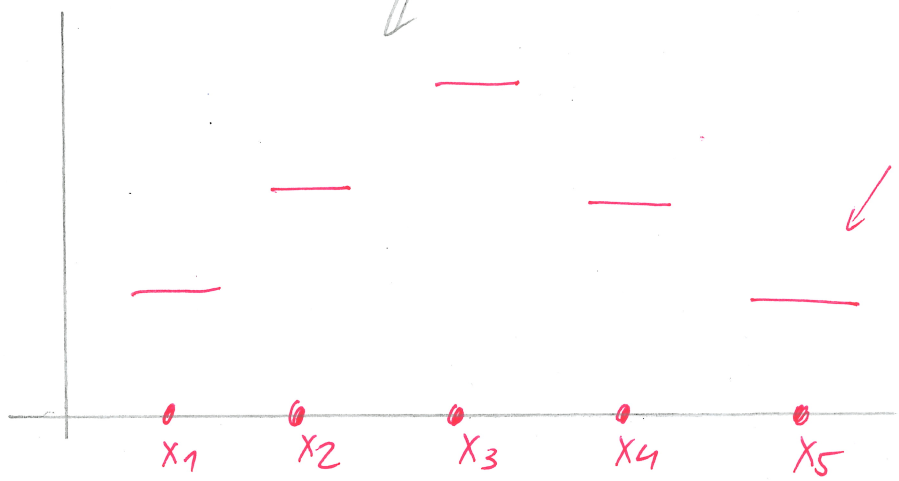



'slide tries to formalize the following idea

we assume if $X^j$ numeric and ~~only~~ included in LM

if only representatives used but still assumed to be linear

more general (model $M_g$)

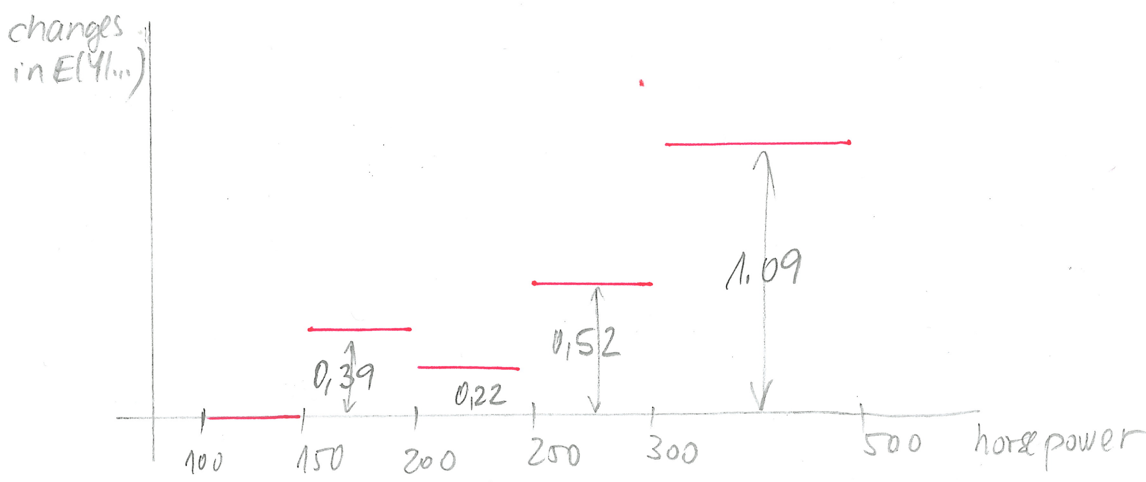



possible fitted values

$\equiv X^j$ categorical with $(G=5)$ levels

**Example:** Test of linearity of effect of horsepower $\equiv X^j$ 



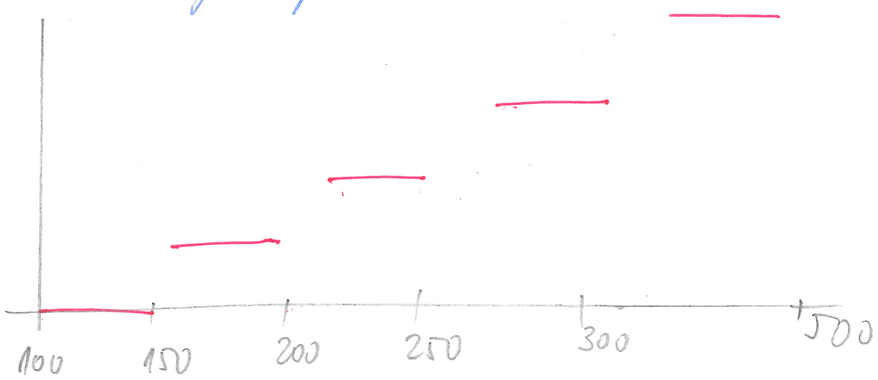


## F-test on submodel

31

horsemid ≡ numeric regressor but only taking values $\in \{125, 175, 225, 275, 400\}$

- use of horsemid as a numeric regressor imposes some structure (linear change) on "group means"




---

32

Approximate F-test : "submodel" uses original (& continuous) values of horsepower

↓

"submodel" is strictly speaking not submodel

---

## Drawback of tests for linearity of the effect

33

- Linearity of the effect of the $j^{th}$ regressor ≡ null hypothesis

→ linearity can be rejected but never confirmed