Least-squares regression

Contents

What changes

The model

Case study

Sensitivity analysis

Exercises

Learning objectives (1/2)

At the end of this chapter you will be able to:

- write least squares as a QP;
- read the KKT conditions as the normal equations;

Learning objectives (2/2)

At the end of this chapter you will be able to:

- compare least squares, absolute deviations and quantiles on the same data;
- shrink the coefficients with a quadratic penalty (ridge).

Prerequisites

- From the module *LP Modelling*, Chapter 10 «Robust and quantile regression: estimating parameters with an LP»: regression with absolute deviations and quantiles as an LP, on the same data
- From this module, Chapter 2 «Background: convexity, quadratic programming and KKT conditions»: convex quadratic programs, the KKT conditions and the multipliers as shadow prices
- From the module *Linear Algebra*, Chapter 6 «System of linear equations»: linear systems, which give the normal equations

What changes

The case and the new objective

- Case, data and linear model: LP Modelling, "Robust and quantile regression"
- There: sum of the **absolute** deviations, an LP
- Here: sum of the **squared** deviations, least squares, a convex QP
- The comparison shows what is gained and lost by changing the objective

The model

QP model of least squares

QP model of least squares

$$\min \quad \sum_{i=1}^{n} r_i^2 \tag{13.1a}$$

$$\text{subject to} \quad \sum_{j=1}^{p} w_j x_{ij} + b + r_i = y_i, \quad \forall i \in \{1, \ldots, n\}, \tag{13.1b}$$

$$w_j \gtreqless 0, \quad \forall j \in \{1, \ldots, p\}, \tag{13.1c}$$

$$b \gtreqless 0, \tag{13.1d}$$

$$r_i \gtreqless 0, \quad \forall i \in \{1, \ldots, n\}. \tag{13.1e}$$

- One free residual r_i per observation: the square does not distinguish the sign

Without features: the mean

Instance

Seven sales: 88, 96, 104, 112, 120, 132, 148 items; $p = 0$

- Step 1: $\tilde{b} = 800/7 = 114.2857$, $\tilde{z} = 2619.4286$, $\tilde{\pi}_i = 2\tilde{r}_i$
- Step 2: $2r_i = \pi_i$ and $\sum_i \pi_i = 0 \Rightarrow \sum_i r_i = 0$: the **mean**
- Step 3: median (LP, $\tau = 1/2$) 112; 9/13 quantile: 120
- Step 4: 148 $\rightarrow$ 248: the mean rises by 14.29, the median stays at 112

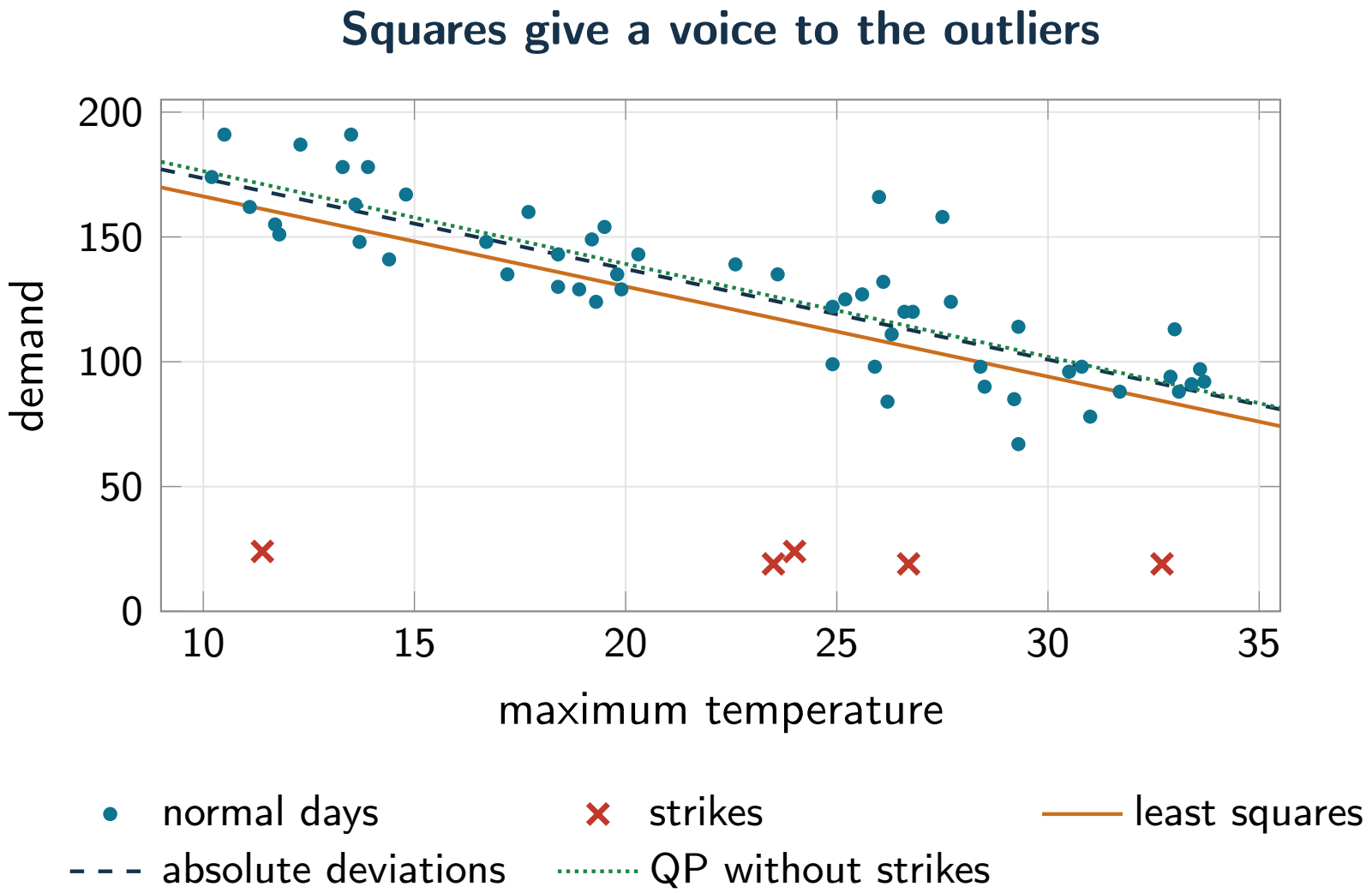
The KKT conditions are the normal equations

$$\pi_i = 2r_i, \qquad \sum_{i=1}^{n} r_i = 0, \qquad \sum_{i=1}^{n} r_i\, x_{ij} = 0 \quad \forall j$$

- Residuals summing to zero and orthogonal to every feature
- The multiplier is twice the residual: a distant datum weighs as much as it is distant
- In the LP the duals were $\pm 1/2$ or an interior value

Case study

Least squares against absolute deviations



The numbers

```
QP (least squares)        : demand = 202.34 - 3.61 temp
LP (absolute deviations): demand = 209.77 - 3.63 temp
QP without strikes        : demand = 213.58 - 3.72 temp
Share of squares due to the strikes: 71.4%
```

- Five days out of sixty make 71% of the objective
- Least squares must come with a check for outliers

Mean plus stock or quantile

```
Typical day (28 degrees, price 3.00, weekend):
  forecast mean (QP)                    128.7 items
  mean + z*sigma, z = 0.5024            133.5 items
  forecast 9/13 quantile (LP)           132.1 items
```

- Mean plus $z\sigma$ assumes normal residuals
- Quantile regression estimates the quantile with no assumption on the errors

SAPIENZA UNIVERSITÀ DI ROMA

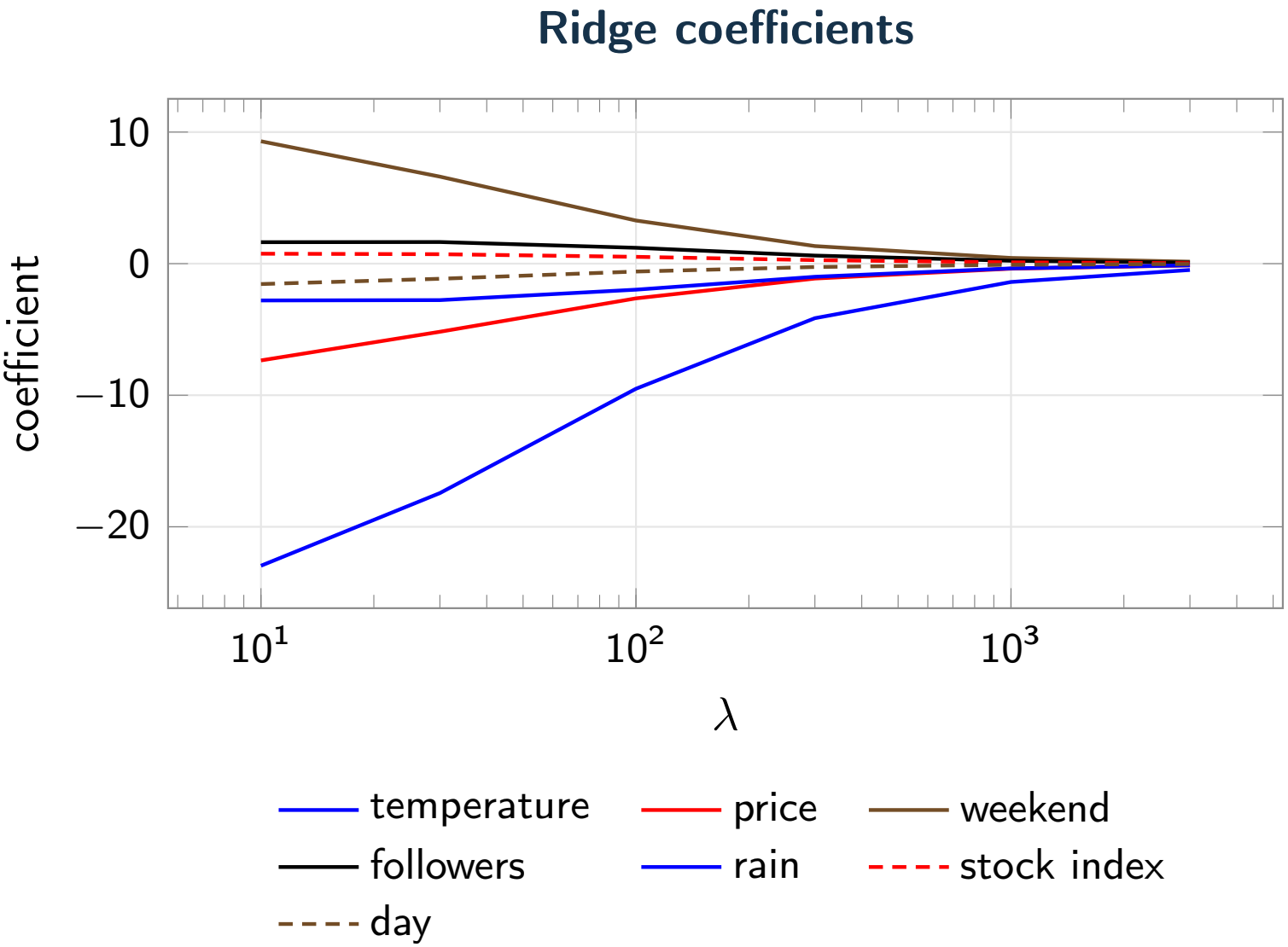
Sensitivity analysis

A single outlying day

shift	QP: b	w	LP: b	w
0	213.58	-3.723	210.49	-3.578
-100	220.59	-4.113	211.63	-3.689
-200	227.60	-4.503	211.63	-3.689

- The QP moves in proportion to the error; the LP changes once and then stops
- A hot day, far from the centre of the data, has a high **leverage**

Ridge regression



- $\lambda \sum_j w_j^2$: the coefficients shrink, but none becomes zero

Exercises

Exercises

1. **By hand**: sum of squares for $b = 112$ and $b = 800/7$
2. **Normal equations**: solve them on the 55 normal days
3. **Weights**: weight 0.01 on the strike days
4. **Huber**: quadratic loss up to 20, linear beyond
5. **Validation**: estimate on the first 40 days, test on the others

Summary

- Least squares is a convex QP; the KKT conditions are the normal equations
- It estimates the mean: excellent with symmetric errors, fragile with outliers
- For the newsvendor: mean plus stock, or the quantile directly
- Ridge stabilizes, the linear budget selects