Searching for Extreme Cases with Regression

Lutz Dümbgen (UniBE), Werner Stahel (ETHZ)

Zürich, August 2018

Setting

Increasing compulsion to assess and rank quality in many domains.

Setting

Increasing compulsion to assess and rank quality in many domains.

Task: Assess and compare the performance of n units.

Raw measures of the units' performance (success or malfunction):

$$Y_1, Y_2, \ldots, Y_n$$

Raw measures of the units' performance (success or malfunction):

$$Y_1, Y_2, \dots, Y_n$$

Examples:

- General practitioners $\quad Y_i =$ mean expenditures per case

Raw measures of the units' performance (success or malfunction):

$$Y_1, Y_2, \ldots, Y_n$$

Examples:

- General practitioners — $Y_i =$ mean expenditures per case
- Psychiatric hospitals — $Y_i =$ mean success per case ($\text{HoNOS}_{\text{in}} - \text{HoNOS}_{\text{out}}$)

Raw measures of the units' performance (success or malfunction):

$$Y_1, Y_2, \ldots, Y_n$$

Examples:

- General practitioners — $Y_i =$ mean expenditures per case
- Psychiatric hospitals — $Y_i =$ mean success per case ($\mathrm{HoNOS}_{\mathrm{in}} - \mathrm{HoNOS}_{\mathrm{out}}$)
- Nursing homes — $Y_i =$ percentage of patients with decubitus

Raw measures of the units' performance (success or malfunction):

$$Y_1, Y_2, \ldots, Y_n$$

Examples:

- General practitioners — $Y_i =$ mean expenditures per case
- Psychiatric hospitals — $Y_i =$ mean success per case ($\text{HoNOS}_{\text{in}} - \text{HoNOS}_{\text{out}}$)
- Nursing homes — $Y_i =$ percentage of patients with decubitus
- Postal delivery areas — $Y_i =$ mean delivery time per item

Raw measures of the units' performance (success or malfunction):

$$Y_1, Y_2, \ldots, Y_n$$

Examples:

- General practitioners — $Y_i =$ mean expenditures per case
- Psychiatric hospitals — $Y_i =$ mean success per case ($\text{HoNOS}_{\text{in}} - \text{HoNOS}_{\text{out}}$)
- Nursing homes — $Y_i =$ percentage of patients with decubitus
- Postal delivery areas — $Y_i =$ mean delivery time per item
- Social welfare offices (Canton of Berne)
  $Y_i =$ expenditure for social welfare per *inhabitant* of municipality

Problem: Units have to work under different frame conditions
(which they cannot influence).

Comparing units in terms of Y_i means
comparing apples and oranges.

For unit i:

$$X_i \; = \big(X_i(1), X_i(2), \ldots, X_i(d)\big) \; \in \; \mathcal{X}$$

(tuple of covariables describing its frame conditions)

For unit i:

$$X_i = \big(X_i(1), X_i(2), \ldots, X_i(d)\big) \in \mathcal{X}$$

(tuple of covariables describing its frame conditions)

Example of social welfare services (municipalities):

- $Y_i =$ expenditures for social welfare per inhabitant

For unit i:

$$X_i \ = \big(X_i(1), X_i(2), \ldots, X_i(d)\big) \ \in \ \mathcal{X}$$

(tuple of covariables describing its frame conditions)

Example of social welfare services (municipalities):

- Y_i = expenditures for social welfare per inhabitant
- $X_i(1)$ = percentage of foreigners

For unit i:

$$X_i = \big(X_i(1), X_i(2), \ldots, X_i(d)\big) \in \mathcal{X}$$

(tuple of covariables describing its frame conditions)

Example of social welfare services (municipalities):

- Y_i = expenditures for social welfare per inhabitant
- $X_i(1)$ = percentage of foreigners
- $X_i(2)$ = percentage of refugees

For unit i:

$$X_i = \big(X_i(1), X_i(2), \ldots, X_i(d)\big) \in \mathcal{X}$$

(tuple of covariables describing its frame conditions)

Example of social welfare services (municipalities):

- Y_i = expenditures for social welfare per inhabitant
- $X_i(1)$ = percentage of foreigners
- $X_i(2)$ = percentage of refugees
- $X_i(3)$ = percentage of people getting subsidies (Ergänzungsleistungen)

For unit i:

$$X_i = \big(X_i(1), X_i(2), \ldots, X_i(d)\big) \in \mathcal{X}$$

(tuple of covariables describing its frame conditions)

Example of social welfare services (municipalities):

- $Y_i =$ expenditures for social welfare per inhabitant
- $X_i(1) =$ percentage of foreigners
- $X_i(2) =$ percentage of refugees
- $X_i(3) =$ percentage of people getting subsidies (Ergänzungsleistungen)
- $X_i(4) =$ percentage of vacant appartments

For unit i:

$$X_i = \big(X_i(1), X_i(2), \ldots, X_i(d)\big) \in \mathcal{X}$$

(tuple of covariables describing its frame conditions)

Example of social welfare services (municipalities):

- Y_i = expenditures for social welfare per inhabitant
- $X_i(1)$ = percentage of foreigners
- $X_i(2)$ = percentage of refugees
- $X_i(3)$ = percentage of people getting subsidies (Ergänzungsleistungen)
- $X_i(4)$ = percentage of vacant appartments
- $X_i(5)$ = percentage of single parents
- ...

Common approach

Regression model:

$$Y_i \;=\; f_o(X_i) + \pi_i$$

with

Common approach

Regression model:

$$Y_i \;=\; f_o(X_i) + \pi_i$$

with

- unknown regression function $f_o : \mathcal{X} \to \mathbb{R}$,
  $f_o(x) =$ "average performance" under frame conditions $x \in \mathcal{X}$,

Common approach

Regression model:

$$Y_i \; = \; f_o(X_i) + \pi_i$$

with

- unknown regression function $f_o : \mathcal{X} \to \mathbb{R}$,
  $f_o(x) =$ "average performance" under frame conditions $x \in \mathcal{X}$,
- "adjusted performances" $\pi_1, \pi_2, \ldots, \pi_n$.

If we knew f_o, we could reconstruct

$$\pi_i = Y_i - f_o(X_i).$$

If we knew f_o, we could reconstruct

$$\pi_i \;=\; Y_i - f_o(X_i).$$

Proposal: For a given linear model $\mathcal{F}$,

$$f_o \quad \rightsquigarrow \quad \widehat{f} \;\in\; \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} (Y_i - f(X_i))^2,$$

$$\pi_i \quad \rightsquigarrow \quad \widehat{\pi}_i \;:=\; Y_i - \widehat{f}(X_i).$$

Traditional regression:

$$
\begin{aligned}
f_o \; &: \; \text{optimal prediction/approximation of } Y \text{ by } X, \\
&\quad\; \text{object of main interest,} \\
\pi_i \; &: \; \text{random noise, approximation errors, nuisance, } \ldots
\end{aligned}
$$

Traditional regression:

f_o : optimal prediction/approximation of Y by X,
object of main interest,

π_i : random noise, approximation errors, nuisance, ...

Present application:

f_o : nuisance part,

π_i : parameters of main interest,
noise neglected ...

Vector notation:

$$\boldsymbol{Y} = f_o(\boldsymbol{X}) + \boldsymbol{\pi}$$

Vector notation:

$$\boldsymbol{Y} \;=\; f_o(\boldsymbol{X}) + \boldsymbol{\pi}$$

Problem: Approach works if, and only if,

Vector notation:

$$\boldsymbol{Y} \;=\; f_o(\boldsymbol{X}) + \boldsymbol{\pi}$$

Problem: Approach works if, and only if,

- linear model is correct, i.e.

$$f_o(\boldsymbol{X}) \;\in\; \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$$

Vector notation:

$$\boldsymbol{Y} \;=\; f_o(\boldsymbol{X}) + \boldsymbol{\pi}$$

Problem: Approach works if, and only if,

- linear model is correct, i.e.

$$f_o(\boldsymbol{X}) \;\in\; \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$$

- no association between adjusted performances and model space, i.e.

$$\boldsymbol{\pi} \;\perp\; \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$$

Assumption

$$\pi \perp \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$$

is questionable.

Assumption

$$\pi \perp \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$$

is questionable.

Example of social welfare services (municipalities):

- Y_i = expenditures for social welfare per inhabitant

Assumption

$$\pi \perp \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$$

is questionable.

Example of social welfare services (municipalities):

- Y_i = expenditures for social welfare per inhabitant
- $X_i(1)$ = percentage of foreigners
  $\rightsquigarrow$ investing in language courses and education

Assumption

$$\pi \perp \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$$

is questionable.

Example of social welfare services (municipalities):

- Y_i = expenditures for social welfare per inhabitant
- $X_i(1)$ = percentage of foreigners
  $\rightsquigarrow$ investing in language courses and education
- $X_i(5)$ = percentage of single parents
  $\rightsquigarrow$ investing in child care and infrastructure for parents
- ...

Even in case of $f_o(\boldsymbol{X}) \in \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$,

$$\begin{aligned} \boldsymbol{Y} &= f_o(\boldsymbol{X}) + \boldsymbol{\pi} \\ &= \widehat{f}(\boldsymbol{X}) + \widehat{\boldsymbol{\pi}} \\ &= \overbrace{f_o(\boldsymbol{X}) + Q\boldsymbol{\pi}} + \overline{Q}\boldsymbol{\pi} \end{aligned}$$

with

$$\begin{aligned} Q &:= \text{orth. projection onto } \{f(\boldsymbol{X}) : f \in \mathcal{F}\}, \\ \overline{Q} &:= I - Q. \end{aligned}$$

More generally, let

$$\boldsymbol{Y} = f_o(\boldsymbol{X}) + \boldsymbol{\pi} + \boldsymbol{\epsilon}$$

with random noise vector $\boldsymbol{\epsilon}$.

More generally, let

$$\boldsymbol{Y} = f_o(\boldsymbol{X}) + \boldsymbol{\pi} + \boldsymbol{\epsilon}$$

with random noise vector $\boldsymbol{\epsilon}$.

Then

$$\begin{aligned}
\boldsymbol{Y} &= f_o(\boldsymbol{X}) + \boldsymbol{\pi} + \boldsymbol{\epsilon} \\
&= \widehat{f}(\boldsymbol{X}) + \widehat{\boldsymbol{\pi}} \\
&= \overbrace{Qf_o(\boldsymbol{X}) + Q\boldsymbol{\pi} + Q\boldsymbol{\epsilon}} + \overbrace{\overline{Q}f_o(\boldsymbol{X}) + \overline{Q}\boldsymbol{\pi} + \overline{Q}\boldsymbol{\epsilon}}\,.
\end{aligned}$$

Specific example of social welfare offices

Cantonal government installed a bonus-malus system:

Specific example of social welfare offices

Cantonal government installed a bonus-malus system:

$$\begin{cases} \widehat{\pi}_i > +0.3 \cdot \widehat{f}(X_i) & \Longrightarrow \text{ malus payment } 0.1 \cdot E_i \cdot \widehat{\pi}_i \\ \widehat{\pi}_i < -0.3 \cdot \widehat{f}(X_i) & \Longrightarrow \text{ bonus payment } 0.1 \cdot E_i \cdot |\widehat{\pi}_i| \\ |\widehat{\pi}_i| \leq 0.3 \cdot \widehat{f}(X_i) & \Longrightarrow \text{ no payment} \end{cases}$$

with

$$E_i := \text{ population size.}$$

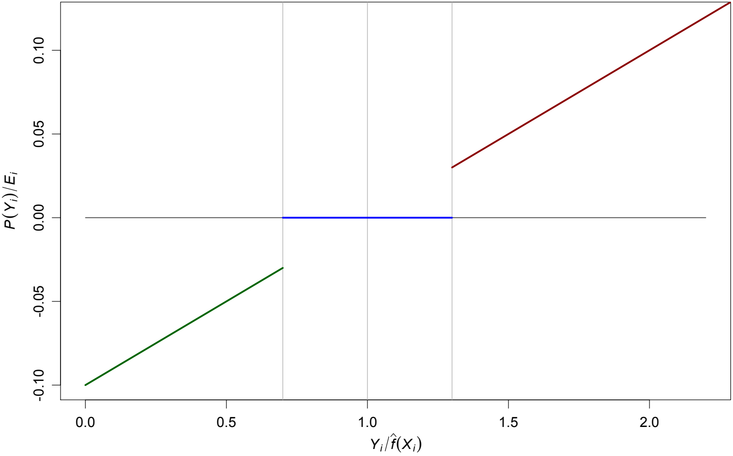
Major flaws of this system

Major flaws of this system

- Questionable assumption $\pi \perp \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$

Major flaws of this system

- Questionable assumption $\pi \perp \{f(\boldsymbol{X}) : f \in \mathcal{F}\}$
- Discontinuous payment function:



- Canton used overly simplistic linear model:

$$\begin{aligned}\widehat{f}(X_i) \;=\; & -146.5 + 10.5 \cdot X_i(1) + \;\; 112.4 \cdot X_i(2) \\ & + 64.9 \cdot X_i(3) + 3850.4 \cdot X_i(4)\end{aligned}$$

- Canton used overly simplistic linear model:

$$\widehat{f}(X_i) = -146.5 + 10.5 \cdot X_i(1) + 112.4 \cdot X_i(2) + 64.9 \cdot X_i(3) + 3850.4 \cdot X_i(4)$$

Adding interactions would lead to “significantly better” fit and different conclusions:

		interactions included		
		malus	no paym.	bonus
no interactions		9	49	8
malus	12	9	**3**	0
no paym.	44	0	43	**1**
bonus	10	0	**3**	7

- Regression function $\widehat{f}$ based on data of one particular year. No specification how to update $\widehat{f}$ in later years!

- Regression function $\widehat{f}$ based on data of one particular year.

  No specification how to update $\widehat{f}$ in later years!

- Covariable $X_i(4) =$ percentage of vacant appartments
  notoriously difficult to measure precisely,

  municipalities may choose between different methods proposed by BfS.

- Naive belief in power of regression method to use

$$Y_i = \text{expenditures per inhabitant}$$

rather than, for instance,

$$Y_i = \text{expenditures per case and year.}$$

- ...

- Naive belief in power of regression method to use

$$Y_i \;=\; \text{expenditures per inhabitant}$$

rather than, for instance,

$$Y_i \;=\; \text{expenditures per case and year.}$$

- ...

Eventually the canton's ordinance had to be cancelled ...

Conclusion

- Regression method potentially useful to identify potentially extreme units

- Alternative approach with case-to-case data:

- Alternative approach with case-to-case data:

  For unit i and case $j \in \{1, \ldots, J_i\}$,

$$Y_{ij} \;=\; f_o(X_{ij}) + \pi_i + \epsilon_{ij}$$

  ...

  Simultaneous inference about the π_i ...

- Alternative approach with case-to-case data:

  For unit i and case $j \in \{1, \ldots, J_i\}$,

$$Y_{ij} \;=\; f_o(X_{ij}) + \pi_i + \epsilon_{ij}$$

  ...

  Simultaneous inference about the π_i ...

- Alternative approach with case-to-case data:

  For unit i and case $j \in \{1, \ldots, J_i\}$,

$$Y_{ij} \;=\; f_o(X_{ij}) + \pi_i + \epsilon_{ij}$$

  . . .

  Simultaneous inference about the π_i . . .

Examples:

- Units: psychiatric hospitals, cases: patients (treatments)
  Y_{ij}: success of treatment
  X_{ij}: features of case (diagnosis, severity, . . .) and
  frame conditions of unit

- Alternative approach with case-to-case data:

  For unit i and case $j \in \{1, \ldots, J_i\}$,

$$Y_{ij} \;=\; f_o(X_{ij}) + \pi_i + \epsilon_{ij}$$

  . . .

  Simultaneous inference about the π_i . . .

Examples:

- Units: psychiatric hospitals, cases: patients (treatments)
  Y_{ij}: success of treatment
  X_{ij}: features of case (diagnosis, severity, . . .) and
  frame conditions of unit
- Units: high schools, cases: pupils
  Y_{ij}: score in 1st year at ETH
  X_{ij}: features of pupils (gender, age, type of matura/school, . . .)