data-analysis

Meng Ye

2024-08-24

## Introduction

The current version re-run the codes of the data analysis.

# environment setup to run ordered logit properly  
options(contrasts = rep("contr.treatment", 2))

### Load packages

library(tidyverse) # package for data cleaning and plotting  
library(readxl)  
library(modelsummary)  
library(ordinal) # package for ordinal logit regression  
library(brant) # brant test for the parallel assumption for ordered logit  
library(MASS) # models that work with the brant test  
library(broom) # extracting model summary as data frame  
library(modelsummary) # deriving model tables  
library(scales) # label percent  
library(lubridate) # working with dates  
library(marginaleffects) #to calculate marginal effects  
library(gt) # to format tables  
library(here) # work with directory  
set.seed(5432)

## Merging CQC and financial data

# import location level full data  
rating<- read\_csv(here("cleaned\_data","cic\_all\_ratings\_2019.csv"))  
finance <- read\_csv(here("cleaned\_data","cls\_finance.csv"))

finance1 <- finance %>%   
 mutate(id\_digit = as.numeric(str\_extract(project\_id, "\\d+"))) %>%  
 arrange(id\_digit)

# checking the largest number for the project\_id in the finance data set   
summary(finance1$id\_digit)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.0 228.0 460.0 610.1 913.2 5181.0

#merging the data  
cic2019 <- rating %>%   
 left\_join(finance1, by = "project\_id")

## data cleanning

#select relevant columns, rename and relabel   
cic\_cleaned <- cic2019 %>%   
 # recode legal form types to be more readable / easier to present  
 mutate(# inherited = ifelse(inherited == "Y", TRUE, FALSE),  
 rating = recode(rating,   
 "Insufficient evidence to rate" = "NA",  
 "Requires improvement" = "Req improv"),  
 date = ymd(publication\_date)) %>%   
   
 # assign order in the rating levels  
 mutate(rating = ordered(rating, levels = c("Inadequate","Req improv", "Good", "Outstanding")),  
 social\_care = ifelse(type == "Social Care Org", "social care", "healthcare")) %>%   
   
   
 # creating a new dummy variable for facility category  
 mutate(founded = as.numeric(founded),  
 year = year(date),  
 age = year - founded,   
 Year = factor(year)) %>%   
 mutate(cls = ifelse(CLS == 1, "CLS", "CLG"),  
 totalequity = as.numeric(totalequity),  
 totalequity\_std = scale(totalequity, center = TRUE, scale = TRUE))

## Warning: There was 1 warning in `mutate()`.  
## ℹ In argument: `founded = as.numeric(founded)`.  
## Caused by warning:  
## ! NAs introduced by coercion

datasummary\_crosstab(  
 formula = cls ~ rating,  
 data = cic\_cleaned  
)

| cls |  | Inadequate | Req improv | Good | Outstanding | All |
| --- | --- | --- | --- | --- | --- | --- |
| CLG | N | 13 | 50 | 412 | 52 | 527 |
|  | % row | 2.5 | 9.5 | 78.2 | 9.9 | 100.0 |
| CLS | N | 44 | 98 | 1078 | 134 | 1354 |
|  | % row | 3.2 | 7.2 | 79.6 | 9.9 | 100.0 |
| All | N | 57 | 151 | 1502 | 186 | 1896 |
|  | % row | 3.0 | 8.0 | 79.2 | 9.8 | 100.0 |

datasummary\_crosstab(  
 formula = cls \* spinout ~ rating,  
 data = cic\_cleaned  
)

| cls | spinout |  | Inadequate | Req improv | Good | Outstanding | All |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CLG | 0 | N | 10 | 27 | 254 | 32 | 323 |
|  |  | % row | 3.1 | 8.4 | 78.6 | 9.9 | 100.0 |
|  | 1 | N | 3 | 23 | 158 | 20 | 204 |
|  |  | % row | 1.5 | 11.3 | 77.5 | 9.8 | 100.0 |
| CLS | 0 | N | 25 | 46 | 643 | 92 | 806 |
|  |  | % row | 3.1 | 5.7 | 79.8 | 11.4 | 100.0 |
|  | 1 | N | 19 | 52 | 435 | 42 | 548 |
|  |  | % row | 3.5 | 9.5 | 79.4 | 7.7 | 100.0 |
|  | All | N | 57 | 151 | 1502 | 186 | 1896 |
|  |  | % row | 3.0 | 8.0 | 79.2 | 9.8 | 100.0 |

## regression analysis

### models without equity variable

model\_order\_overall <- clm(rating ~ cls + spinout + social\_care + age + dissolved,  
 data = filter(cic\_cleaned, domain == "Overall"),  
 link = "logit")  
  
model\_order\_safe <- clm(rating ~ cls + spinout + social\_care + age + dissolved,  
 data = filter(cic\_cleaned, domain == "Safe"),  
 link = "logit")  
model\_order\_effective <- clm(rating ~ cls + spinout + social\_care + age + dissolved,  
 data = filter(cic\_cleaned, domain == "Effective"),  
 link = "logit")  
model\_order\_caring <- clm(rating ~ cls + spinout + social\_care + age + dissolved,  
 data = filter(cic\_cleaned, domain == "Caring"),  
 link = "logit")  
model\_order\_well\_led <- clm(rating ~ cls + spinout + social\_care + age + dissolved,  
 data = filter(cic\_cleaned, domain == "Well-led"),  
 link = "logit")  
model\_order\_responsive <- clm(rating ~ cls + spinout + social\_care + age + dissolved,  
 data = filter(cic\_cleaned, domain == "Responsive"),  
 link = "logit")

ordinal\_models <-  
 modelsummary(  
 list(  
 "overall" = model\_order\_overall,  
 "safe" = model\_order\_safe,  
 "effective" = model\_order\_effective,  
 "caring" = model\_order\_caring,  
 "well-led" = model\_order\_well\_led,  
 "responsive" = model\_order\_responsive  
 ),  
 coef\_omit = "region",  
 exponentiate = F,  
 statistic = "({p.value}) {stars}")  
ordinal\_models

|  | overall | safe | effective | caring | well-led | responsive |
| --- | --- | --- | --- | --- | --- | --- |
| Inadequate|Req improv | -3.485 | -3.363 | -3.775 |  | -3.510 |  |
|  | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* |  | (<0.001) \*\*\* |  |
| Req improv|Good | -2.496 | -1.536 | -2.316 | -5.337 | -2.021 | -3.910 |
|  | (<0.001) \*\*\* | (0.003) \*\* | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* |
| Good|Outstanding | 1.435 | 4.103 | 2.803 | 0.674 | 1.885 | 1.485 |
|  | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* | (0.202) | (<0.001) \*\*\* | (0.007) \*\* |
| clsCLS | 0.483 | 0.564 | 0.273 | -0.329 | -0.056 | -0.135 |
|  | (0.072) + | (0.102) | (0.451) | (0.364) | (0.857) | (0.730) |
| spinout | -0.386 | -0.305 | -0.446 | -0.189 | -0.325 | -0.681 |
|  | (0.064) + | (0.370) | (0.214) | (0.588) | (0.277) | (0.082) + |
| social\_caresocial care | 0.096 | 0.947 | 0.153 | -0.817 | 0.043 | -0.093 |
|  | (0.716) | (0.016) \* | (0.677) | (0.040) \* | (0.888) | (0.810) |
| age | -0.119 | -0.045 | -0.050 | -0.070 | -0.023 | -0.064 |
|  | (<0.001) \*\*\* | (0.354) | (0.275) | (0.207) | (0.587) | (0.247) |
| dissolved | 0.212 | -0.972 | 0.370 | 0.129 | 0.740 | 0.439 |
|  | (0.607) | (0.032) \* | (0.530) | (0.817) | (0.131) | (0.458) |
| Num.Obs. | 540 | 261 | 261 | 261 | 261 | 261 |
| AIC | 906.6 | 340.0 | 332.5 | 274.7 | 445.5 | 269.3 |
| BIC | 940.9 | 368.5 | 361.0 | 299.6 | 474.1 | 294.3 |
| RMSE | 2.45 | 2.17 | 2.23 | 1.55 | 2.42 | 1.40 |

### models with equity variables

Due to the large range/dispersion of the fiancial data. I standardize the totalequity variable to enable the models to run.

summary(cic\_cleaned$totalequity)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -23436 53934 653284 2892894 3613364 36881694 57

eq\_order\_overall <- clm(rating ~ cls + spinout + social\_care + age + dissolved + totalequity\_std,  
 data = filter(cic\_cleaned, domain == "Overall"),  
 link = "logit")  
  
eq\_order\_safe <- clm(rating ~ cls + spinout + social\_care + age + dissolved + totalequity\_std,  
 data = filter(cic\_cleaned, domain == "Safe"),  
 link = "logit")  
eq\_order\_effective <- clm(rating ~ cls + spinout + social\_care + age + dissolved + totalequity\_std,  
 data = filter(cic\_cleaned, domain == "Effective"),  
 link = "logit")  
eq\_order\_caring <- clm(rating ~ cls + spinout + social\_care + age + dissolved + totalequity\_std,  
 data = filter(cic\_cleaned, domain == "Caring"),  
 link = "logit")  
eq\_order\_well\_led <- clm(rating ~ cls + spinout + social\_care + age + dissolved + totalequity\_std,  
 data = filter(cic\_cleaned, domain == "Well-led"),  
 link = "logit")  
eq\_order\_responsive <- clm(rating ~ cls + spinout + social\_care + age + dissolved + totalequity\_std,  
 data = filter(cic\_cleaned, domain == "Responsive"),  
 link = "logit")

eq\_models <-  
 modelsummary(  
 list(  
 "overall" = eq\_order\_overall,  
 "safe" = eq\_order\_safe,  
 "effective" = eq\_order\_effective,  
 "caring" = eq\_order\_caring,  
 "well-led" = eq\_order\_well\_led,  
 "responsive" = eq\_order\_responsive  
 ),  
 coef\_omit = "region",  
 exponentiate = F,  
 statistic = "({p.value}) {stars}")  
eq\_models

|  | overall | safe | effective | caring | well-led | responsive |
| --- | --- | --- | --- | --- | --- | --- |
| Inadequate|Req improv | -3.632 | -3.792 | -3.868 |  | -3.763 |  |
|  | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* |  | (<0.001) \*\*\* |  |
| Req improv|Good | -2.670 | -1.992 | -2.404 | -5.353 | -2.328 | -4.106 |
|  | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* |
| Good|Outstanding | 1.308 | 3.868 | 2.713 | 0.651 | 1.652 | 1.316 |
|  | (<0.001) \*\*\* | (<0.001) \*\*\* | (<0.001) \*\*\* | (0.232) | (<0.001) \*\*\* | (0.019) \* |
| clsCLS | 0.495 | 0.611 | 0.246 | -0.311 | -0.028 | -0.151 |
|  | (0.066) + | (0.083) + | (0.496) | (0.392) | (0.929) | (0.698) |
| spinout | -0.507 | -0.682 | -0.545 | -0.191 | -0.530 | -0.858 |
|  | (0.026) \* | (0.066) + | (0.150) | (0.602) | (0.099) + | (0.041) \* |
| social\_caresocial care | 0.081 | 0.960 | 0.184 | -0.842 | -0.004 | -0.074 |
|  | (0.760) | (0.017) \* | (0.619) | (0.035) \* | (0.989) | (0.849) |
| age | -0.132 | -0.078 | -0.057 | -0.072 | -0.047 | -0.079 |
|  | (<0.001) \*\*\* | (0.094) + | (0.219) | (0.204) | (0.294) | (0.154) |
| dissolved | 0.240 | -0.857 | 0.449 | 0.104 | 0.812 | 0.541 |
|  | (0.566) | (0.066) + | (0.450) | (0.853) | (0.104) | (0.364) |
| totalequity\_std | 0.092 | 0.446 | 0.163 | -0.038 | 0.174 | 0.238 |
|  | (0.356) | (0.027) \* | (0.382) | (0.851) | (0.274) | (0.219) |
| Num.Obs. | 538 | 259 | 259 | 259 | 259 | 259 |
| AIC | 895.3 | 327.6 | 333.0 | 275.6 | 436.8 | 269.2 |
| BIC | 933.8 | 359.6 | 365.1 | 304.1 | 468.8 | 297.7 |
| RMSE | 2.45 | 2.16 | 2.23 | 1.55 | 2.41 | 1.40 |