

《国际关系定量分析基础》2020 秋季

第四次作业 (共计 100 分)

赵佳鹏

2017010454

截止时间：2020 年 11 月 30 日 11: 59 am

注意事项：

- 作业在网络学堂提交
- 请将 Chunk 中的 `eval=FALSE` 改为 `eval=TRUE` 再 `knit`
- 请将文件解压缩后，直接在 R Markdown 文件中完成本次作业
- 学生可以互相讨论作业，但作业必须是自己本人独立完成
- 提交作业的文件名需以 `HW-4-YourName.Rmd`, `HW-4-YourName.pdf` 或者 `HW-4-YourName.html`, 请将 `YourName` 替换为你的姓名。(若 R Markdown 出现无法 `knit` 为 pdf 情况，则使用 `bookdown::html_document2`：会生成为 html)
- 请显示每道题的 R Code 于 pdf 中，注重 Code 的整洁性和可读性，可参考 Google's R Style Guide

本次作业需要的数据已经提供，请将数据与 `HW-4-YourName.Rmd` 放在同一工作路径的文件夹内

```
load("HW-4.RData")
```

本次作业的 `HW-4.RData` 数据包括 `broz_et_al` 和 `map` 两个数据，其中数据 `broz_et_al` 来自于 J. Lawrence Broz, Zhiwen Zhang 以及 Gaoyang Wang 发表于《国际组织》2020 年第 2 期的复制数据（见 J. Lawrence Broz, Zhiwen Zhang, and Gaoyang Wang, “Explaining Foreign Support for China’s Global Economic Leadership,” *International Organization*, Vol. 74, Summer 2020, pp. 417–52）。该文章检验了世界各国对中国一带一路峰会态度的影响因素。

其中部分变量如下：

- `countryname`: The Correlates of War (COW) country name
- `attendance`: “DV=Attendance” (1= Yes; 0 otherwise)
- `obor_nations`: “One Belt, One Road Position” (1= Yes; 0 otherwise)
- `ftas`: “FTA with China” (1= Yes; 0 otherwise)

- bits: “BIT with China” (1= Yes; 0 otherwise)
- fc_dummy_cumcount_bank_s1_9016: “Financial Crises”
- ka_open_sd9016: “Variability of Capital Account Policy”
- mean_portfolio_vol: “Volatility of Portfolio Outflows”
- imf_dummy_unrest_index_sum_9017: “Social Unrest During IMF Programs”
- wto_cases_cumulcount9516: “WTO Complaints Against The U.S.”
- imf_governance_deficit_usd_2015: “IMF Governance Deficit”

表-1 统计了部分变量的统计分布特征。请利用 `broz_et_al` 数据完成以下各题。

表 1: 变量的描述性统计

Statistic	N	Mean	Median	Max	Min	St. Dev.
attendance	192	0.151	0	1	0	0.359
obor_nations	192	0.349	0	1	0	0.478
ftas	192	0.125	0	1	0	0.332
bits	192	0.547	1	1	0	0.499
fc_dummy_cumcount_bank_s1_9016	162	4.358	0.000	27.000	0.000	6.475
ka_open_sd9016	178	0.140	0.116	0.431	0.000	0.105
mean_portfolio_vol	93	0.162	0.011	7.324	0.00002	0.797
imf_dummy_unrest_index_sum_9017	192	13.625	2	188	0	26.984
wto_cases_cumulcount9516	192	0.651	0	17	0	2.246
imf_governance_deficit_usd_2015	184	0.066	0.052	1.924	−7.653	0.630

数据可视化

1.(20 分) 利用 `map` 数据和 `ggplot` 绘制一幅各国参加一带一路峰会的地图（提示：参与国家的变量为 `attendance`）。根据地图，你发现参与国家的地理分布有何模式和特征？

```
ggplot(data = map) +
# 设置图形横纵坐标与填充属性
  geom_polygon(aes(x = long, y = lat,
                    group = group, fill = attendance)) +
  coord_fixed() +
  scale_fill_manual(values = c("blue", "red"),
                    na.value = "gray") +
```

```
theme(line = element_blank(),  
      legend.position = "right",  
      legend.title=element_blank(),  
      panel.border=element_blank(),  
      panel.grid=element_blank(),  
      axis.ticks=element_blank(),  
      axis.text=element_blank()) +  
# 更改默认横纵坐标名使之更明确  
labs(y = "latitude", x = "longitude")
```

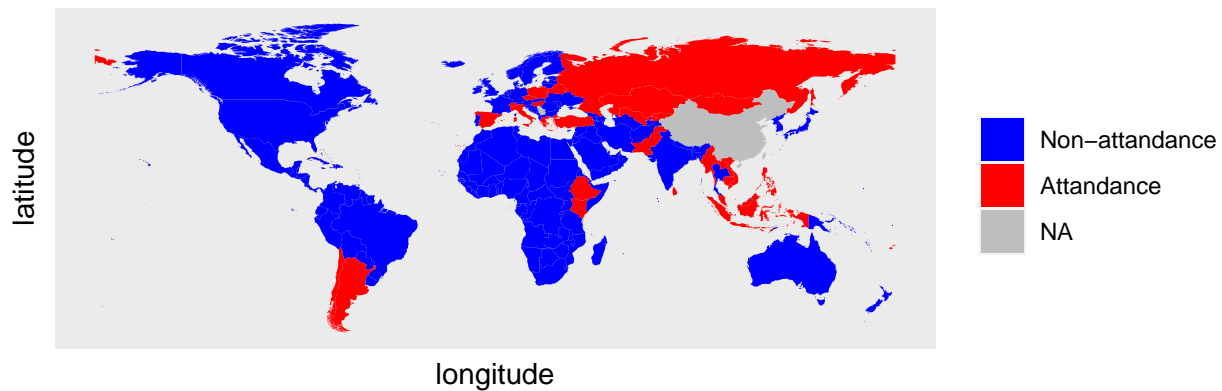


图 1: Countries attended BRI Submit

根据地图，我们发现参与国家地理分布不均，主要分布在以下地区：

- 南美洲南部（如阿根廷、智利）；
- 东南亚（如印度尼西亚、菲律宾）；
- 中东欧（如白俄罗斯、俄罗斯）；
- 地中海北岸（如西班牙、意大利）；

- 东非（如埃塞俄比亚、肯尼亚）。

其它地区亦有少量参与一带一路峰会的国家，如中亚的哈萨克斯坦、南亚的巴基斯坦、大洋洲的斐济等。大部分参与一带一路峰会的国家都是沿海国家。

估计二分类因变量模型

2.（15 分）利用线性概率模型（linear probability model）估计以下模型，将其命名为 `m1`，并制作一个回归表格。根据回归表格，解读对应的回归系数 β_1 。

LPM:

$$attendance = \beta_0 + \beta_1 * \text{One Belt, One Road Position} + \beta_2 * \text{FTA with China} + \beta_3 * \text{BIT with China} + \epsilon$$

```
m1 <- lm(attendance ~ obor_nations + ftas + bits, data = broz_et_al)
library(stargazer)
# 报告回归结果，并设定置信区间水平为 95%
stargazer(m1, single.row = TRUE, ci = TRUE, ci.level=0.95, type = "latex",
  header = FALSE, title = " 线性概率模型回归统计结果")
```

表 2: 线性概率模型回归统计结果

	Dependent variable:
	attendance
obor_nations	0.129** (0.021, 0.238)
ftas	0.304*** (0.159, 0.450)
bits	0.126** (0.026, 0.226)
Constant	−0.001 (−0.070, 0.068)
Observations	192
R ²	0.212
Adjusted R ²	0.200
Residual Std. Error	0.321 (df = 188)
F Statistic	16.906*** (df = 3; 188)

Note: *p<0.1; **p<0.05; ***p<0.01

本回归模型中，自变量 `obor_nations` 的回归系数 β_1 为 0.129。它在 95% 置信水平上是显著的。这意味着在此水平下，在其它变量不变的情况下，位于被中国划定为优先进行基础设施投资的商路沿线的国家相比于不位于其沿线的国家而言，参加一带一路峰会的概率平均高出 12.9%。

3. (15 分) 利用 Logit 回归估计以下模型，将其命名为 `m2`，并制作一个回归表格。根据回归表格，解读对应的回归系数 β_1 和 β_2 。

Logit Model:

$$P(\text{attendance} = 1) = \beta_0 + \beta_1 * \text{Financial Crises} + \\ \beta_2 * \text{Variability of Capital Account Policy} + \\ \beta_3 * \text{Volatility of Portfolio Outflows} + \\ \beta_4 * \text{Social Unrest During IMF Programs} + \\ \beta_5 * \text{WTO Complaints Against The U.S.} + \\ \beta_6 * \text{IMF Governance Deficit} + \epsilon$$

```
m2 <- glm(attendance ~ fc_dummy_cumcount_bank_s1_9016 +
           ka_open_sd9016 + mean_portfolio_vol +
           imf_dummy_unrest_index_sum_9017 +
           wto_cases_cumulcount9516 + imf_governance_deficit_usd_2015,
           data = broz_et_al, family = binomial(link = "logit"))
# 报告回归结果，并设定置信区间水平为 95%
stargazer(m2, single.row = TRUE, ci = TRUE, ci.level=0.95, type = "latex",
           header = FALSE, title = "Logit 模型回归统计结果")
```

本回归模型中，自变量 `fc_dummy_cumcount_bank_s1_9016`、`ka_open_sd9016` 的回归系数 β_1 、 β_2 分别为 0.037、2.044，均大于 0。这意味着在此水平下，在其它变量不变的情况下：

- 一国在历史上每多一次金融危机，它参加一带一路峰会概率的 log-odds 值平均增加 0.037。这将会使参加一带一路峰会的概率上升。
- 一国的标准化 Chinn-Ito 指数的标准差每增加 1，它参加一带一路峰会概率的 log-odds 值平均增加 2.044。这将会使参加一带一路峰会的概率上升。

然而，它们在 90% 置信水平上都不显著。这影响了模型的可靠性。

注：自变量 `fc_dummy_cumcount_bank_s1_9016` 不是二分变量，它统计了 1990 年至 2016 年间一国发生重大金融危机的总次数，参见论文原文。

4. (20 分) 利用线性概率模型 (linear probability model)、Logit 和 Probit 分别估计以下模型，将其命名为 `m3`、`m4`、`m5`，同时制作一个回归系数图。

表 3: Logit 模型回归统计结果

	<i>Dependent variable:</i>
	attendance
fc_dummy_cumcount_bank_s1_9016	0.037 (−0.057, 0.131)
ka_open_sd9016	2.044 (−3.020, 7.108)
mean_portfolio_vol	0.517 (−0.314, 1.347)
imf_dummy_unrest_index_sum_9017	0.014 (−0.006, 0.033)
wto_cases_cumulcount9516	−0.109 (−0.368, 0.149)
imf_governance_deficit_usd_2015	0.144 (−0.825, 1.114)
Constant	−2.023*** (−3.192, −0.855)
Observations	85
Log Likelihood	−41.836
Akaike Inf. Crit.	97.671

Note:

*p<0.1; **p<0.05; ***p<0.01

$$\begin{aligned}
P(\text{attendance} = 1) = & \beta_0 + \beta_1 * \text{One Belt, One Road Position} + \\
& \beta_2 * \text{FTA with China} + \\
& \beta_3 * \text{BIT with China} + \\
& \beta_4 * \text{Financial Crises} + \\
& \beta_5 * \text{Variability of Capital Account Policy} + \\
& \beta_6 * \text{Volatility of Portfolio Outflows} + \\
& \beta_7 * \text{Social Unrest During IMF Programs} + \\
& \beta_8 * \text{WTO Complaints Against The U.S.} + \\
& \beta_9 * \text{IMF Governance Deficit} + \epsilon
\end{aligned}$$

#LPM

```

m3 <- lm(attendance ~ obor_nations + ftas + bits +
         fc_dummy_cumcount_bank_s1_9016 +
         ka_open_sd9016 + mean_portfolio_vol +
         imf_dummy_unrest_index_sum_9017 +
         wto_cases_cumulcount9516 + imf_governance_deficit_usd_2015,

```

```
data = broz_et_al)

#probit
m4 <- glm(attendance ~ obor_nations + ftas + bits +
          fc_dummy_cumcount_bank_s1_9016 +
          ka_open_sd9016 + mean_portfolio_vol +
          imf_dummy_unrest_index_sum_9017 +
          wto_cases_cumulcount9516 + imf_governance_deficit_usd_2015,
          data = broz_et_al, family = binomial(link = "probit"))

#logit
m5 <- glm(attendance ~ obor_nations + ftas + bits +
          fc_dummy_cumcount_bank_s1_9016 +
          ka_open_sd9016 + mean_portfolio_vol +
          imf_dummy_unrest_index_sum_9017 +
          wto_cases_cumulcount9516 + imf_governance_deficit_usd_2015,
          data = broz_et_al, family = binomial(link = "logit"))

library(dotwhisker)
# 作图报告回归系数，并设定置信区间水平为 95%
dwplot(list(m3, m4, m5), conf.level = .95, show_intercept = TRUE) +
  theme_bw() +
  ggtitle("Coefficient Plot")
```

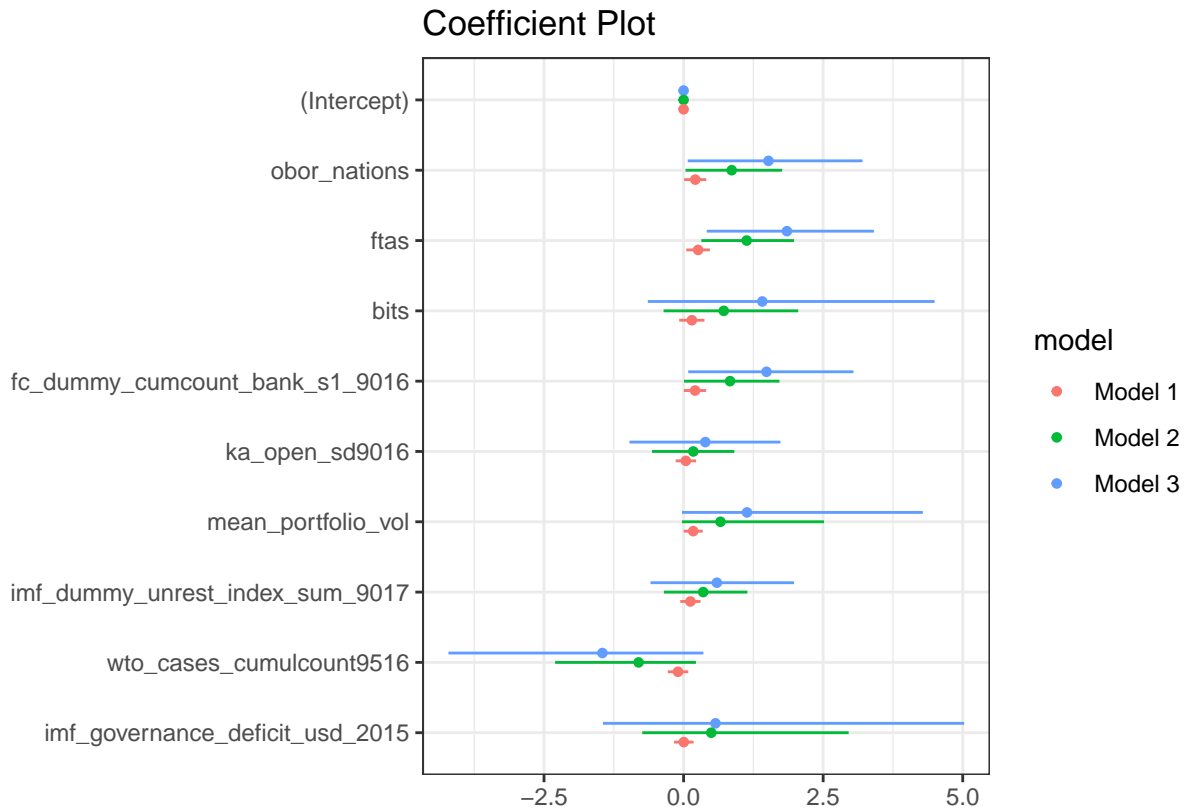


图 2: Coefficient Plots

5. (30 分) 根据 `broz_et_al` 数据和 Logit 模型 5, 计算澳大利亚 (Australia)、法国 (France) 和印度 (India) 支持中国经济领导的概率 (包括 95% 置信区间)。(提示: 根据模型 5 结果以及这三国对应变量取值, 计算预测概率及其置信区间)

```
au_newdata <- data.frame(broz_et_al) %>%
  filter(countryname == "Australia") %>%
  select(obor_nations:imf_governance_deficit_usd_2015)
fr_newdata <- data.frame(broz_et_al) %>%
  filter(countryname == "France") %>%
  select(obor_nations:imf_governance_deficit_usd_2015)
ind_newdata <- data.frame(broz_et_al) %>%
  filter(countryname == "India") %>%
  select(obor_nations:imf_governance_deficit_usd_2015)
# 定义计算置信区间、建立数据框的函数
calc <- function(x) {data.frame(Predicted_possibility = x,
  Upper_value = x + (1.96*x),
```


表 4: 三国支持中国经济领导的概率

Country	Predicted_possibility	Upper_value	Lower_value
Australia	0.215	0.636	-0.206
France	0.028	0.083	-0.027
India	0.333	0.987	-0.320

```

Lower_value = x - (1.96*x))}
au_prob <- predict(m5, au_newdata, type = "response", se.fit = TRUE)
au_df <- calc(au_prob$fit)
fr_prob <- predict(m5, fr_newdata, type = "response", se.fit = TRUE)
fr_df <- calc(fr_prob$fit)
ind_prob <- predict(m5, ind_newdata, type = "response", se.fit = TRUE)
ind_df <- calc(ind_prob$fit)
# 创建表格并输出
df <- bind_rows(au_df, fr_df, ind_df) %>%
  data.frame(Country = c("Australia", "France", "India"), .)
kable(df, digits = 3, row.names = FALSE,
      caption = " 三国支持中国经济领导的概率")

```