BRSM Project

Dhruv Hirpara, Hariharan Kalimuthu, Kishore Kumar Srujana Vanka

06/05/2023

Contents

Introduction

Data Analysis

Importing Libraries and Data Preprocessing

```
# import libraries
library(readxl)
library(ggplot2)
library(car)
library(reshape2)
library(haven)
library(dplyr)
library(zoo)
library(metafor)
library(tidyr)
library(forestplot)
```

```
formatify <- function(df, code) {
  df <- df[df$Code == code, c("Year", tail(names(df), 1))]
  colnames(df)[ncol(df)] <- "value"
  return(df)
}

print_info <- function(df) {
  cat(deparse(substitute(df)), "\n")
  col1 <- df[,1]
  cat("[", min(col1), ", ", max(col1), "]", "\n")
  head(df)
}

adjusted_net_savings <- read.csv("data/adjusted-net-saving-current-us-dollars.csv")
adjusted_net_savings <- formatify(adjusted_net_savings, "ISR")
```

```r
annual_co2_emissions <- read.csv("data/annual-co2-emissions-per-country.csv")
annual_co2_emissions <- formatify(annual_co2_emissions, "ISR")

child_mortality <- read.csv("data/child-mortality-around-the-world.csv")
child_mortality <- formatify(child_mortality, "ISR")

gdp_per_capita <- read.csv("data/gdp-per-capita-worldbank.csv")
gdp_per_capita <- formatify(gdp_per_capita, "ISR")

government_revenue_sharegdp <- read.csv("data/government-revenue-as-a-share-of-gdp.csv")
government_revenue_sharegdp <- formatify(government_revenue_sharegdp, "ISR")

GDP <- read.csv("data/gross-domestic-product.csv")
GDP <- formatify(GDP, "ISR")

human_development_index <- read.csv("data/human-development-index.csv")
human_development_index <- formatify(human_development_index, "ISR")

labor_productivity <- read.csv("data/labor-productivity-per-hour-PennWorldTable.csv")
labor_productivity <- formatify(labor_productivity, "ISR")

life_expectancy <- read.csv("data/life-expectancy.csv")
life_expectancy <- formatify(life_expectancy, "ISR")

research_spending_sharegdp <- read.csv("data/research-spending-gdp.csv")
research_spending_sharegdp <- formatify(research_spending_sharegdp, "ISR")

solar_electricity_percapita <- read.csv("data/solar-electricity-per-capita.csv")
solar_electricity_percapita <- formatify(solar_electricity_percapita, "ISR")

unemployment_rate <- read.csv("data/unemployment-rate.csv")
unemployment_rate <- formatify(unemployment_rate, "ISR")

education_in_govt_exp <- read.csv("data/share-of-education-in-government-expenditure.csv")
education_in_govt_exp <- formatify(education_in_govt_exp, "ISR")

labour_data <- read.csv("romer-model-data/Labor_Data.csv")
labour_data <- subset(labour_data, Country.Code == "ISR")
labour_data <- labour_data[, 5:ncol(labour_data)]
col_names <- names(labour_data)
new_col_names <- substring(col_names, 2)
names(labour_data) <- new_col_names

labour_data <- melt(labour_data)
labour_data <- labour_data[, !colSums(is.na(labour_data)) == nrow(labour_data)]
labour_data <- na.omit(labour_data)
colnames(labour_data) <- c("Year", "value")
head(labour_data)
```

```
##   Year   value
## 1 1990 1900701
## 2 1991 2035534
## 3 1992 2133756
## 4 1993 2252545
```

```
## 5 1994 2353183
## 6 1995 2442219
```

```
print_info(adjusted_net_savings)
```

```
## adjusted_net_savings
## [ 1970 ,  2019 ]
```

```
##      Year      value
## 2295 1970  724273088
## 2296 1971 1260111232
## 2297 1972 2017276672
## 2298 1973 2126510976
## 2299 1974 1793309056
## 2300 1975 1352494464
```

```
print_info(annual_co2_emissions)
```

```
## annual_co2_emissions
## [ 1930 ,  2021 ]
```

```
##       Year value
## 14394 1930 39977
## 14395 1931 39977
## 14396 1932 50880
## 14397 1933 65417
## 14398 1934 69051
## 14399 1935 90857
```

```
print_info(child_mortality)
```

```
## child_mortality
## [ 1950 ,  2021 ]
```

```
##      Year   value
## 7561 1950 4.29607
## 7562 1951 4.31432
## 7563 1952 4.32142
## 7564 1953 4.26406
## 7565 1954 4.20126
## 7566 1955 4.13682
```

```
print_info(gdp_per_capita)
```

```
## gdp_per_capita
## [ 1995 ,  2020 ]
```

```
##      Year    value
## 2604 1995 26233.13
## 2605 1996 27096.08
## 2606 1997 27464.76
## 2607 1998 27970.64
## 2608 1999 28240.99
## 2609 2000 29950.32
```

```
print_info(government_revenue_sharegdp)
```

```
## government_revenue_sharegdp
## [ 2000 ,  2020 ]
```

```
##      Year value
## 1344 2000 43.76
## 1345 2001 43.54
## 1346 2002 43.25
## 1347 2003 41.84
## 1348 2004 40.91
## 1349 2005 41.00
```

```
print_info(GDP)
```

```
## GDP
## [ 1995 ,  2020 ]
```

```
##      Year        value
## 4599 1995 139622940672
## 4600 1996 148039122944
## 4601 1997 153849528320
## 4602 1998 160307806208
## 4603 1999 166031736832
## 4604 2000 180795719680
```

```
print_info(human_development_index)
```

```
## human_development_index
## [ 1990 ,  2021 ]
```

```
##      Year value
## 2544 1990 0.787
## 2545 1991 0.793
## 2546 1992 0.797
## 2547 1993 0.804
## 2548 1994 0.808
## 2549 1995 0.811
```

```
print_info(labor_productivity)
```

```
## labor_productivity
## [ 1981 ,  2019 ]
```

```
##      Year    value
## 1577 1981 25.19910
## 1578 1982 25.77843
## 1579 1983 26.18268
## 1580 1984 26.19285
## 1581 1985 26.88302
## 1582 1986 28.37297
```

```
print_info(life_expectancy)
```

```
## life_expectancy
## [ 1950 ,  2021 ]
```

```
##      Year value
## 8486 1950  68.2
## 8487 1951  68.1
## 8488 1952  68.0
## 8489 1953  68.1
## 8490 1954  68.3
## 8491 1955  68.7
```

```
print_info(research_spending_sharegdp)
```

```
## research_spending_sharegdp
## [ 1996 ,  2018 ]
```

```
##     Year   value
## 932 1996 2.59054
## 933 1997 2.80736
## 934 1998 2.91897
## 935 1999 3.33008
## 936 2000 3.93329
## 937 2001 4.18499
```

```
print_info(solar_electricity_percapita)
```

```
## solar_electricity_percapita
## [ 1965 ,  2021 ]
```

```
##      Year value
## 3423 1965     0
## 3424 1966     0
## 3425 1967     0
## 3426 1968     0
## 3427 1969     0
## 3428 1970     0
```

```
print_info(unemployment_rate)
```

```
## unemployment_rate
## [ 1991 ,  2021 ]
```

```
##      Year value
## 2636 1991 13.39
## 2637 1992 14.08
## 2638 1993 12.74
## 2639 1994  9.93
## 2640 1995  8.78
## 2641 1996  8.46
```

```
library(dplyr)
library(tidyr)
library(purrr)

# Create a list of dataframes to merge
dfs <- list(adjusted_net_savings, annual_co2_emissions, child_mortality, gdp_per_capita,
            government_revenue_sharegdp, GDP, human_development_index, labor_productivity,
            life_expectancy, research_spending_sharegdp, solar_electricity_percapita, unemployment_rate
            education_in_govt_exp)

# Get the names of the original dataframes
df_names <- c("Year", "adjusted_net_savings", "annual_co2_emissions", "child_mortality", "gdp_per_capita
            government_revenue_sharegdp", "GDP", "human_development_index", "labor_productivity", "
            life_expectancy", "research_spending_sharegdp", "solar_electricity_percapita", "unemployment
            "education_in_govt_exp")

# Merge dataframes using the "Year" column as the key and keep the original dataframe names
merged_df <- Reduce(function(x, y) {
```

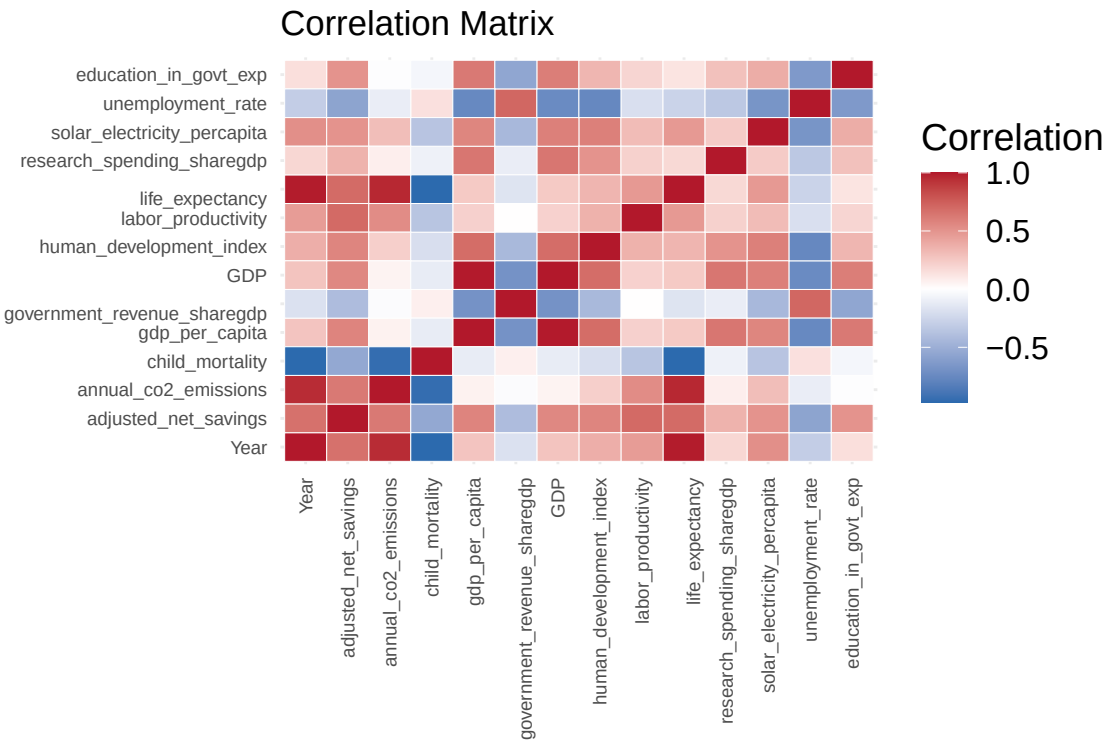
```
  suffix_x <- df_names[which(df_names == deparse(substitute(x)))][1]
  suffix_y <- df_names[which(df_names == deparse(substitute(y)))][1]
  merge(x, y, by = "Year", all = TRUE, suffixes = c(paste0(".", suffix_x), paste0(".", suffix_y)))
}, dfs)
merged_df <- set_names(merged_df, df_names)

# Remove rows where Year <= 1960
merged_df <- subset(merged_df, Year > 1960)

# Fill missing values using mean imputation
for (col in names(merged_df)[-1]) {
  merged_df[is.na(merged_df[, col]), col] <- mean(merged_df[, col], na.rm = TRUE)
}
```

```
# Compute correlation matrix
cor_matrix <- cor(merged_df, use = "complete.obs")

# Plot correlation matrix
melted_matrix <- melt(cor_matrix)
melted_matrix <- melted_matrix %>%
  arrange(desc(value))
ggplot(melted_matrix, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "#2166ac", mid = "white", high = "#b2182b",
                       midpoint = 0, na.value = "gray") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5, size = 6.5),
    axis.text.y = element_text(angle = 0, hjust = 1, vjust = 0.5, size = 6.5),
    axis.title.x = element_text(size = 14),
    axis.title.y = element_text(size = 14),
    legend.text = element_text(size = 12),
    legend.title = element_text(size = 14)
  ) +
  labs(title = "Correlation Matrix", x = "", y = "", fill = "Correlation")
```




```
lm2 <- lm(GDP ~ Year + labor_productivity, data = merged_df)
summary(lm2)
```

```
##
## Call:
## lm(formula = GDP ~ Year + labor_productivity, data = merged_df)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -9.574e+10 -1.153e+10  8.796e+09  1.484e+10  9.498e+10
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -9.017e+11  6.646e+11  -1.357    0.180
## Year                5.497e+08  3.458e+08   1.590    0.117
## labor_productivity  1.220e+09  1.602e+09   0.762    0.449
##
## Residual standard error: 4.216e+10 on 58 degrees of freedom
## Multiple R-squared:  0.08485,    Adjusted R-squared:  0.05329
## F-statistic: 2.689 on 2 and 58 DF,  p-value: 0.07643
```

```
coef(lm2)
```

```
##        (Intercept)               Year labor_productivity
##      -901674983520          549734478         1219663665
```

```
# Plot linear regression model
ggplot(merged_df, aes(x = Year, y = GDP, color = labor_productivity)) +
  geom_point() +
  scale_color_gradient(low = "white", high = "blue") +
```

```
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  labs(x = "Labor Productivity", y = "GDP", color = "Research Spending (Share of GDP)") +
  theme_minimal()
```