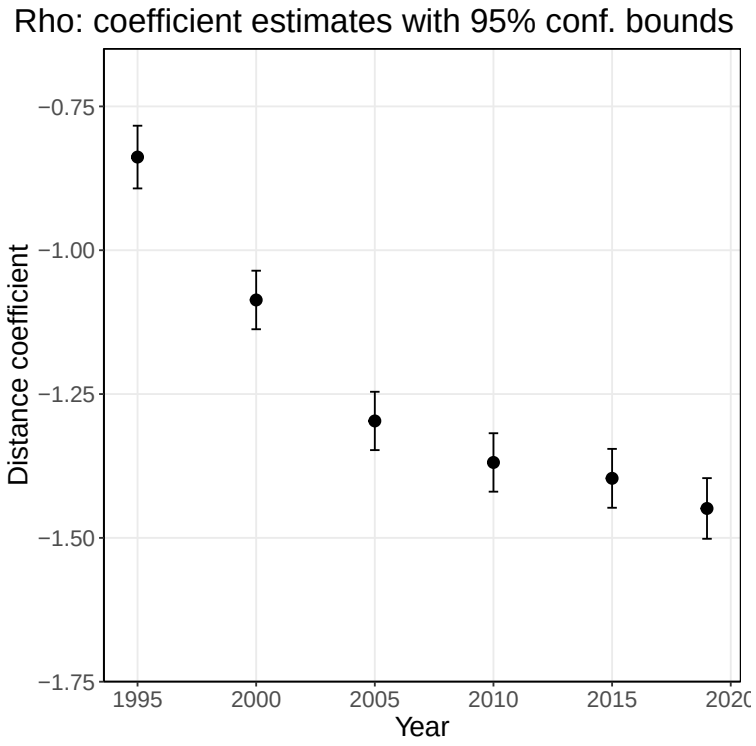
November 24, 2022

The results below are generated from an R script.

```
### packages

library(stargazer)
library(tidyverse)
library(fixest)
library(devtools)
library(gapminder)
library(xtable)
source("functions/collect_coefs.R")

# Set your wd here
setwd("~/Desktop/predoc/adao_kehre_lorenzoni")

theme_1 <- theme_bw() +
  theme(axis.line = element_line(colour = "black"),
        panel.grid.minor = element_blank(),
        panel.border = element_rect(colour = "black"),
        panel.background = element_blank(),
        plot.title = element_text(hjust = 0.5),
        text = element_text(size = 16),
        plot.title.position = "plot")

### Importing data
trade <- read_csv("data/trade.csv")

## Rows:  4450681 Columns:  5
## - Column specification --------------------------------
## Delimiter:  ","
## chr (1):  hs2
## dbl (4):  year, origin, destination, trade
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

gravity <- read_csv("data/gravity.csv")

## Rows:  366054 Columns:  8
## - Column specification --------------------------------
## Delimiter:  ","
## dbl (8):  year, origin, destination, distance, contiguity, language, colonial, rta
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
### Task 2: Data manipulation

## a) collapse

trade_all <- trade %>%
  group_by(year, origin, destination) %>%
  summarize(trade = sum(trade))

## `summarise()` has grouped output by 'year', 'origin'.  You can override using the
## `.groups` argument.

## b) merging

tr_merged <- trade_all %>%
  inner_join(gravity, # inner join will only keep obs present in both dtasets
             by = c("year", "origin", "destination"))

## c) descriptive stats

desc_stat2015 <- tr_merged %>%
  filter(year == 2015) %>%
  ungroup() %>%
  select(-c(year, origin, destination)) %>%
  na.omit() %>%
  gather(Variable, value) %>%
  # Summarize by variable
  group_by(Variable) %>%
  # summarise all columns
  summarise(N = sum(!is.na(value)),
            `Mean` = mean(value),
            `Std. Dev.` = sd(value),
            `Median` = median(value),
            `10th Percentile` = quantile(value, .1),
            `90th Percentile` = quantile(value, .9))

# create a latex table
desc_tab <- xtable(desc_stat2015,
                   digits = 2,
                   caption = 'Summary statistics for year 2015') %>%
  print(type = "latex",
        include.rownames = FALSE,
        format.args = list(big.mark = ","),
        file = "tables/sum2015.tex")

### Task 3: Estimation

## a)

# select 2015 and add logged variables
tr_2015 <- tr_merged %>%
  filter(year == 2015) %>%
  na.omit() %>%
  mutate(log_dist = log(distance),
         log_tr = log(trade))
```

```r
# create bins
nbins = 50

pdf(file = "fig/binplot.pdf")
tr_2015 %>%
  # here I just use equally spaced bins for simplicity
  mutate(bin = ntile(log_tr, n = nbins)) %>%
  group_by(bin) %>%
  # I summarize the variables by taking the mean inside the bins
  summarise(log_tr = mean(log_tr), log_dist = mean(log_dist)) %>%
  ggplot(aes(x = log_tr, y = log_dist)) +
  geom_point() +
  theme_1 +
  labs(title = paste0("Distance and trade: binplot, ", nbins, " bins", sep = ""),
       x = "log(trade), bin mean",
       y = "log(distance), bin mean")
dev.off()

## RStudioGD
##         2

# pearson correlation
print(cor.test(tr_2015$log_tr, tr_2015$log_dist))
```




```
##
##  Pearson's product-moment correlation
##
## data:  tr_2015$log_tr and tr_2015$log_dist
## t = -44.021, df = 26936, p-value < 2.2e-16
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2701682 -0.2478874
## sample estimates:
##        cor
## -0.2590623

## b)

ols_year <- feols(log(trade) ~ log(distance),
                  data = tr_merged, split = ~year)

## NOTE: 5,460 observations removed because of NA values (RHS: 5,460).

# collect coefficients manually
coefs_dst <- collect_coefs(model = ols_year,
                           variable = "log(distance)",
                           confIntr = 0.95)
coefs_dst <- as.data.frame(coefs_dst) %>%
  mutate_if(is.character, as.numeric)


# plot
pdf(file = "fig/dist_coef_simple.pdf")
coefs_dst %>%
  ggplot(aes(y = coef, x = smpl)) +
  geom_point(aes(stroke = 1.5)) +
  geom_errorbar(aes(ymin = right, ymax = left, width = .4)) +
  theme_1 +
  scale_y_continuous(limits = c(-1.7, -0.7)) +
  theme(plot.title = element_text(hjust = 0.8),
        plot.margin = margin(1,1,1.5,1.2, "cm"))  +
  labs(title = TeX("$rho^t$ estimates with 95% conf. bounds"),
       x = "Year",
       y = "Distance coefficient")
dev.off()

## RStudioGD
##         2

## c)

# since we estimate for each year separately, origin and destination FEs
# will suffice instead of origin x year and destination x year
ols_fe <- feols(log(trade) ~ log(distance) | origin + destination,
                data = tr_merged, split = ~year)

## NOTE: 5,460 observations removed because of NA values (RHS: 5,460).

# collect the coefficients
coefs_dst_fe <- collect_coefs(model = ols_fe,
                              variable = "log(distance)",
                              confIntr = 0.95)
coefs_dst_fe <- as.data.frame(coefs_dst_fe) %>%
  mutate_if(is.character, as.numeric)
```

```r
# plot
pdf(file = "fig/dist_coef_fe.pdf")
coefs_dst_fe %>%
  ggplot(aes(y = coef, x = smpl)) +
  geom_point(aes(stroke = 1.5)) +
  geom_errorbar(aes(ymin = right, ymax = left, width = .4)) +
  theme_1 +
  scale_y_continuous(limits = c(-2, -1.5)) +
  theme(plot.title = element_text(hjust = 0.8),
        plot.margin = margin(1, 1, 1.5, 1.2, "cm"))  +
  labs(title = TeX("$beta^t$ estimates with 95% conf. bounds"),
       x = "Year",
       y = "Distance coefficient")
dev.off()

## RStudioGD
##         2

## d)


# estimation
ols_full <- feols(log(trade) ~ log(distance) +
                    contiguity + language + colonial + rta
                  | origin + destination,
                  data = tr_2015)
# table
etable(ols_full, tex = TRUE, file = "tables/full.tex", digits = 3, replace = T)
```

The R session information (including the OS info, R version and all packages used):

```r
sessionInfo()

## R version 4.2.1 (2022-06-23)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Monterey 12.3.1
##
## Matrix products: default
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] latex2exp_0.9.5 xtable_1.8-4    gapminder_0.3.0 binsreg_0.7     devtools_2.4.5
##  [6] usethis_2.1.6   fixest_0.11.0   forcats_0.5.2   stringr_1.4.1   dplyr_1.0.9
## [11] purrr_0.3.4     readr_2.1.2     tidyr_1.2.0     tibble_3.1.8    ggplot2_3.3.6
## [16] tidyverse_1.3.2 stargazer_5.2.3
##
## loaded via a namespace (and not attached):
##  [1] googledrive_2.0.0   colorspace_2.0-3    ellipsis_0.3.2      fs_1.5.2
##  [5] rstudioapi_0.14     farver_2.1.1        MatrixModels_0.5-0  remotes_2.4.2
```

```
##  [9] bit64_4.0.5         fansi_1.0.3         lubridate_1.8.0     xml2_1.3.3
## [13] splines_4.2.1       cachem_1.0.6        knitr_1.39          pkgload_1.3.0
## [17] Formula_1.2-4       jsonlite_1.8.0      broom_1.0.0         dbplyr_2.2.1
## [21] shiny_1.7.2         compiler_4.2.1      httr_1.4.4          backports_1.4.1
## [25] assertthat_0.2.1    Matrix_1.4-1        fastmap_1.1.0       gargle_1.2.0
## [29] cli_3.4.1           later_1.3.0         htmltools_0.5.3     quantreg_5.94
## [33] prettyunits_1.1.1   tools_4.2.1         gtable_0.3.0        glue_1.6.2
## [37] dreamerr_1.2.3      tinytex_0.41        Rcpp_1.0.9          cellranger_1.1.0
## [41] vctrs_0.5.0         nlme_3.1-157        xfun_0.32           ps_1.7.1
## [45] rvest_1.0.3         mime_0.12           miniUI_0.1.1.1      lifecycle_1.0.3
## [49] googlesheets4_1.0.1 MASS_7.3-57         zoo_1.8-10          scales_1.2.1
## [53] vroom_1.5.7         hms_1.1.2           promises_1.2.0.1    parallel_4.2.1
## [57] sandwich_3.0-2      SparseM_1.81        yaml_2.3.5          memoise_2.0.1
## [61] stringi_1.7.8       highr_0.9           pkgbuild_1.3.1      rlang_1.0.6
## [65] pkgconfig_2.0.3     matrixStats_0.63.0  evaluate_0.16       lattice_0.20-45
## [69] htmlwidgets_1.5.4   labeling_0.4.2      bit_4.0.4           processx_3.7.0
## [73] tidyselect_1.1.2    magrittr_2.0.3      R6_2.5.1            generics_0.1.3
## [77] profvis_0.3.7       DBI_1.1.3           pillar_1.8.1        haven_2.5.1
## [81] withr_2.5.0         survival_3.3-1      modelr_0.1.9        crayon_1.5.1
## [85] utf8_1.2.2          tzdb_0.3.0          rmarkdown_2.15      urlchecker_1.0.1
## [89] grid_4.2.1          readxl_1.4.1        callr_3.7.2         reprex_2.0.2
## [93] digest_0.6.29       httpuv_1.6.6        numDeriv_2016.8-1.1 munsell_0.5.0
## [97] sessioninfo_1.2.2
```

```
Sys.time()
```

```
## [1] "2022-11-24 16:37:54 +04"
```