FinalProject

Wenjing Ding/Jiechen Qiu/Sizhe Jia/Zexi Xu

2025-11-30

```
# All used libraries
# RMarkdown & knitting
library(knitr)
library(rmarkdown)

# Data manipulation
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.1     ✔ stringr   1.5.2
## ✔ ggplot2   4.0.0     ✔ tibble    3.3.0
## ✔ lubridate 1.9.4     ✔ tidyr     1.3.1
## ✔ purrr     1.1.0
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

```
library(dplyr)
library(haven)

# Regression
library(fixest)

# Table output
library(modelsummary)
library(gt)
library(ggrepel)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```r
#Part 3

# ---  Load Data ---
ajr_data <- read_dta("~/Desktop/maketable1.dta")

# ---  Prepare Data Samples ---
base_sample <- ajr_data %>% filter(baseco == 1)
quartile_sample <- base_sample %>%
  filter(!is.na(extmort4)) %>%
  mutate(quartile = ntile(extmort4, 4))

# ---  Define Helper Functions to Calculate Stats ---
get_means <- function(df, vars) {
  df %>% summarise(across(all_of(vars), ~mean(.x, na.rm = TRUE)))
}
get_sds <- function(df, vars) {
  df %>% summarise(across(all_of(vars), ~sd(.x, na.rm = TRUE)))
}

# ---  Calculate All Statistics ---

all_vars <- c("logpgp95", "loghjypl", "avexpr", "cons00a",
              "cons1", "democ00a", "euro1900", "logem4")

# Calculate means and sds for all groups
means_w <- get_means(ajr_data, all_vars)
sds_w <- get_sds(ajr_data, all_vars)
means_b <- get_means(base_sample, all_vars)
sds_b <- get_sds(base_sample, all_vars)
means_q <- quartile_sample %>% group_by(quartile) %>% get_means(all_vars)
sds_q <- quartile_sample %>% group_by(quartile) %>% get_sds(all_vars)

# Get observation counts
obs_w <- nrow(ajr_data)
obs_b <- nrow(base_sample)
obs_q <- quartile_sample %>% group_by(quartile) %>% count()

# ---  Assemble the Final Table ---
final_table <- tibble(
  ` ` = c("Log GDP per capita (PPP) in 1995", "(Std. Dev.)",
          "Log output per worker in 1988", "(Std. Dev.)",
          "Average protection against expropriation risk, 1985-1995", "(Std. Dev.)",
          "Constraint on executive in 1900", "(Std. Dev.)",
          "Constraint on executive in first year of independence", "(Std. Dev.)",
          "Democracy in 1900", "(Std. Dev.)",
          "European settlements in 1900", "(Std. Dev.)",
          "Log European settler mortality", "(Std. Dev.)",
          "Number of observations", ""),
  `Whole world` = c(means_w$logpgp95, sds_w$logpgp95, means_w$loghjypl, sds_w$loghjypl, means_w$av
expr, sds_w$avexpr,
                    means_w$cons00a, sds_w$cons00a, means_w$cons1, sds_w$cons1, means_w$democ00a,
sds_w$democ00a,
                    means_w$euro1900, sds_w$euro1900, means_w$logem4, sds_w$logem4, obs_w, NA),
  `Base sample` = c(means_b$logpgp95, sds_b$logpgp95, means_b$loghjypl, sds_b$loghjypl, means_b$av
expr, sds_b$avexpr,
                    means_b$cons00a, sds_b$cons00a, means_b$cons1, sds_b$cons1, means_b$democ00a,
sds_b$democ00a,
                    means_b$euro1900, sds_b$euro1900, means_b$logem4, sds_b$logem4, obs_b, NA),
```

```
  `(1)` = c(means_q$logpgp95[1], sds_q$logpgp95[1], means_q$loghjypl[1], sds_q$loghjypl[1], means_
q$avexpr[1], sds_q$avexpr[1],
            means_q$cons00a[1], sds_q$cons00a[1], means_q$cons1[1], sds_q$cons1[1], means_q$democ0
0a[1], sds_q$democ00a[1],
            means_q$euro1900[1], sds_q$euro1900[1], means_q$logem4[1], sds_q$logem4[1], obs_q$n
[1], NA),
  `(2)` = c(means_q$logpgp95[2], sds_q$logpgp95[2], means_q$loghjypl[2], sds_q$loghjypl[2], means_
q$avexpr[2], sds_q$avexpr[2],
            means_q$cons00a[2], sds_q$cons00a[2], means_q$cons1[2], sds_q$cons1[2], means_q$democ0
0a[2], sds_q$democ00a[2],
            means_q$euro1900[2], sds_q$euro1900[2], means_q$logem4[2], sds_q$logem4[2], obs_q$n
[2], NA),
  `(3)` = c(means_q$logpgp95[3], sds_q$logpgp95[3], means_q$loghjypl[3], sds_q$loghjypl[3], means_
q$avexpr[3], sds_q$avexpr[3],
            means_q$cons00a[3], sds_q$cons00a[3], means_q$cons1[3], sds_q$cons1[3], means_q$democ0
0a[3], sds_q$democ00a[3],
            means_q$euro1900[3], sds_q$euro1900[3], means_q$logem4[3], sds_q$logem4[3], obs_q$n
[3], NA),
  `(4)` = c(means_q$logpgp95[4], sds_q$logpgp95[4], means_q$loghjypl[4], sds_q$loghjypl[4], means_
q$avexpr[4], sds_q$avexpr[4],
            means_q$cons00a[4], sds_q$cons00a[4], means_q$cons1[4], sds_q$cons1[4], means_q$democ0
0a[4], sds_q$democ00a[4],
            means_q$euro1900[4], sds_q$euro1900[4], means_q$logem4[4], sds_q$logem4[4], obs_q$n
[4], NA)
)

# ---  Generate and Style the HTML Table ---
final_table %>%
  # Use a single, robust mutate with case_when to format all numbers
  mutate(across(where(is.numeric), ~ case_when(
    # Condition 0: If the value is NA, make it a blank string first.
    is.na(.) ~ "",

    # Condition 1: Handle the "Number of observations" row
    ` `[row_number()] == "Number of observations" ~ sprintf("%.0f", .),

    # Condition 2: Handle the standard deviation rows
    grepl("Std. Dev.", ` `[row_number()]) ~ sprintf("(%.2f)", .),

    # Condition 3 (Default): Handle all other numeric rows (the means)
    TRUE ~ sprintf("%.2f", .)
  ))) %>%

  # Replace any other NA values with a blank string (mostly for safety)
  mutate(across(everything(), ~ replace_na(., ""))) %>%

  # Generate the kable table
  kable("html", caption = "<b>TABLE 1—DESCRIPTIVE STATISTICS</b>", align = "r") %>%
  kable_styling(bootstrap_options = c("striped", "condensed"), full_width = FALSE) %>%
  add_header_above(c(" " = 3, "By quartiles of mortality" = 4)) %>%
  column_spec(1, extra_css = "padding-left: 2em;") %>%
  row_spec(seq(1, 16, 2), bold = TRUE)
```

TABLE 1—DESCRIPTIVE STATISTICS

			By quartiles of mortality			
	Whole world	**Base sample**	**1.**	**2.**	**3.**	**4.**
Log GDP per capita (PPP) in 1995	**8.30**	**8.06**	**8.84**	**8.39**	**7.83**	**7.18**
(Std. Dev.)	(1.07)	(1.04)	(1.23)	(0.64)	(0.79)	(0.61)
Log output per worker in 1988	**-1.71**	**-1.93**	**-1.06**	**-1.49**	**-2.11**	**-3.05**
(Std. Dev.)	(1.08)	(0.98)	(0.82)	(0.42)	(0.75)	(0.47)
Average protection against expropriation risk, 1985-1995	**6.99**	**6.52**	**7.74**	**6.45**	**6.13**	**5.74**
(Std. Dev.)	(1.83)	(1.47)	(1.51)	(1.01)	(1.19)	(1.39)
Constraint on executive in 1900	**1.85**	**2.25**	**3.86**	**3.06**	**1.13**	**1.00**
(Std. Dev.)	(1.79)	(2.11)	(2.91)	(2.02)	(0.52)	(0.00)
Constraint on executive in first year of independence	**3.63**	**3.40**	**4.57**	**2.62**	**3.27**	**3.27**
(Std. Dev.)	(2.39)	(2.39)	(2.95)	(1.96)	(2.28)	(2.15)
Democracy in 1900	**1.12**	**1.64**	**4.07**	**2.47**	**0.20**	**0.00**
(Std. Dev.)	(2.54)	(3.00)	(4.30)	(2.88)	(0.41)	(0.00)
European settlements in 1900	**30.10**	**16.18**	**31.50**	**25.00**	**8.18**	**1.00**
(Std. Dev.)	(41.86)	(25.53)	(42.22)	(15.60)	(12.19)	(2.73)
Log European settler mortality	**4.61**	**4.66**	**3.16**	**4.29**	**4.88**	**6.29**
(Std. Dev.)	(1.30)	(1.26)	(0.69)	(0.05)	(0.39)	(0.76)
Number of observations	376	64	16	16	16	16

```
## part 4

# ---  Load Data ---
ajr_dta <- read_dta("~/Desktop/maketable2.dta")

# ---  Create Data Subsets ---
base_sample <- ajr_dta %>% filter(baseco == 1)

# ---  Run All Regressions ---
# Note on results: The public data has minor differences from the paper's,
# so coefficients may not match perfectly.
model_list <- list(
  "(1)" = feols(logpgp95 ~ avexpr, data = ajr_dta, se = "hetero"),
  "(2)" = feols(logpgp95 ~ avexpr, data = base_sample, se = "hetero"),
  "(3)" = feols(logpgp95 ~ avexpr + lat_abst, data = ajr_dta, se = "hetero"),
  "(4)" = feols(logpgp95 ~ avexpr + lat_abst + africa + asia + other, data = ajr_dta, se = "heter
o"),
  "(5)" = feols(logpgp95 ~ avexpr + lat_abst, data = base_sample, se = "hetero"),
  "(6)" = feols(logpgp95 ~ avexpr + lat_abst + africa + asia + other, data = base_sample, se = "he
tero"),
  "(7)" = feols(loghjypl ~ avexpr, data = ajr_dta, se = "hetero"),
  "(8)" = feols(loghjypl ~ avexpr, data = base_sample, se = "hetero")
)
```

```
## NOTE: 52 observations removed because of NA values (LHS: 15, RHS: 42).
```

```
## NOTE: 52 observations removed because of NA values (LHS: 15, RHS: 43).
## NOTE: 52 observations removed because of NA values (LHS: 15, RHS: 43).
```

```
## NOTE: 55 observations removed because of NA values (LHS: 40, RHS: 42).
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3).
```

```
# ---  Define Table Components ---
gof_map <- list(
  list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
  list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)
)


coef_map <- c("avexpr"    = "Average Expropriation Risk",
              "lat_abst"  = "Distance from Equator",
              "africa"    = "Africa",
              "asia"      = "Asia",
              "other"     = "Other continents")


modelsummary(
  model_list,
  output = "gt",
  title = "Table 2: OLS Regressions",
  coef_map = coef_map,
  gof_map = gof_map,
  stars = c('*' = .1, '**' = .05, '***' = .01),
  add_rows = tribble(
    ~term, ~"(1)", ~"(2)", ~"(3)", ~"(4)", ~"(5)", ~"(6)", ~"(7)", ~"(8)",
    "Base Sample", "No", "Yes", "No", "No", "Yes", "Yes", "No", "Yes",
    "Continent Dummies", "No", "No", "No", "Yes", "No", "Yes", "No", "No"
  ),
  notes = "Notes: Robust standard errors are in parentheses."
) %>%
  # Use gt's tab_spanner to create the correct headers
  tab_spanner(
    label = "Dependent variable: Log GDP per capita, 1995",
    columns = 2:7 # Selects columns for models (1) through (6)
  ) %>%
  tab_spanner(
    label = "Dependent variable: Log output per worker, 1988",
    columns = 8:9 # Selects columns for models (7) and (8)
  )
```

Table 2: OLS Regressions

	Dependent variable: Log GDP per capita, 1995						Dependent variable: Log output per worker, 1988	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Expropriation Risk	0.532***	0.522***	0.463***	0.390***	0.468***	0.401***	0.446***	0.457***
	(0.029)	(0.050)	(0.052)	(0.051)	(0.063)	(0.064)	(0.029)	(0.050)
Distance from Equator			0.872*	0.333	1.577**	0.875		
			(0.499)	(0.442)	(0.651)	(0.614)		

* p < 0.1, ** p < 0.05, *** p < 0.01

Notes: Robust standard errors are in parentheses.

	Dependent variable: Log GDP per capita, 1995						Dependent variable: Log output per worker, 1988	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Africa				-0.916***		-0.881***		
				(0.154)		(0.156)		
Asia				-0.153		-0.577*		
				(0.181)		(0.299)		
Other continents				0.304*		0.107		
				(0.174)		(0.223)		
Num. Obs.	111	64	111	111	64	64	108	61
R-squared	0.611	0.540	0.623	0.715	0.574	0.714	0.554	0.486
Base Sample	No	Yes	No	No	Yes	Yes	No	Yes
Continent Dummies	No	No	No	Yes	No	Yes	No	No

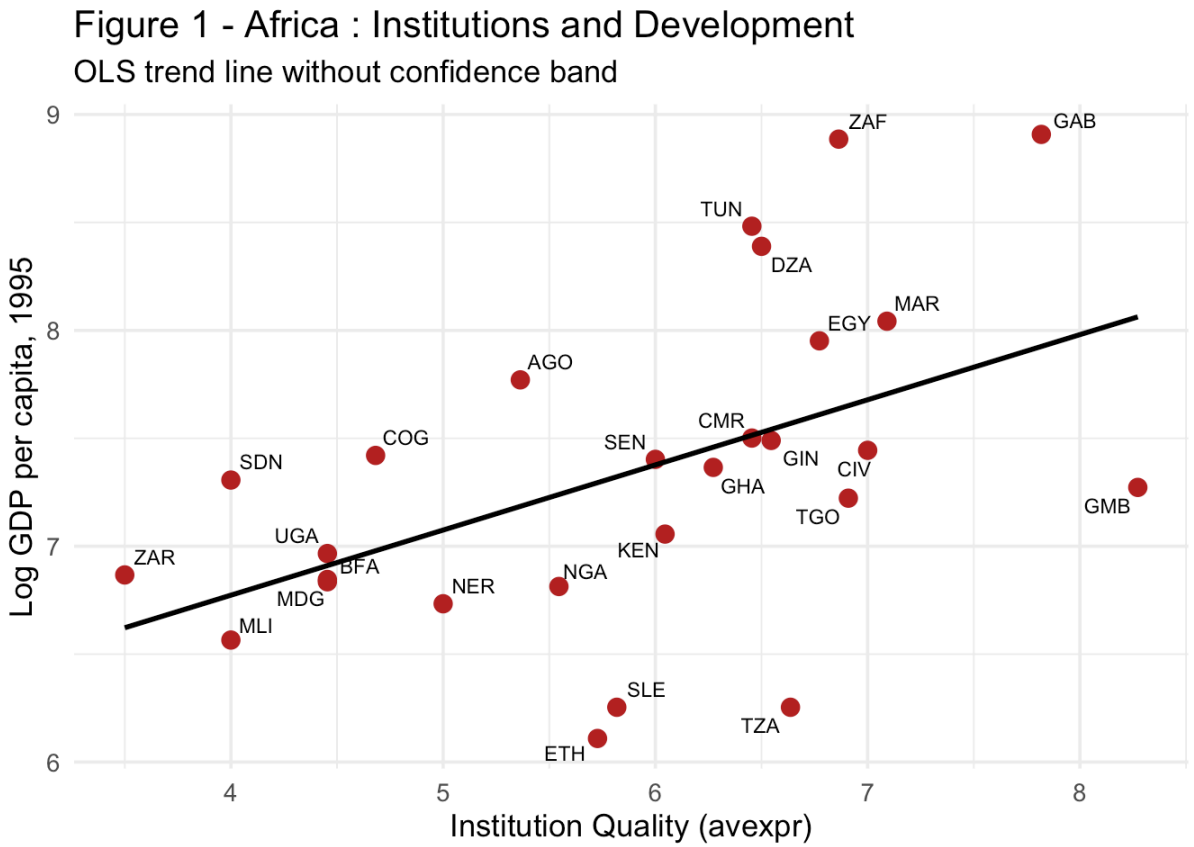
* p < 0.1, ** p < 0.05, *** p < 0.01

Notes: Robust standard errors are in parentheses.

```
base_sample <- base_sample %>%
  mutate(region = case_when(
    africa == 1 ~ "Africa",
    asia == 1 ~ "Asia",
    TRUE ~ "Other"
  ))

region_colors <- c(
  "Africa" = "firebrick",
  "Asia" = "steelblue",
  "Other" = "darkgreen"
)

regions <- unique(base_sample$region)

for (r in regions) {
  p <- ggplot(
    base_sample %>% filter(region == r),
    aes(x = avexpr, y = logpgp95, label = shortnam)
  ) +
    geom_point(size = 3, color = region_colors[r]) +  # Different color each plot
    geom_smooth(method = "lm", se = FALSE, linewidth = 1, color = "black") +
    geom_text_repel(size = 3) +
    theme_minimal(base_size = 13) +
    labs(
      title = paste("Figure 1 –", r, ": Institutions and Development"),
      subtitle = "OLS trend line without confidence band",
      x = "Institution Quality (avexpr)",
      y = "Log GDP per capita, 1995"
    )

  print(p)
}
```

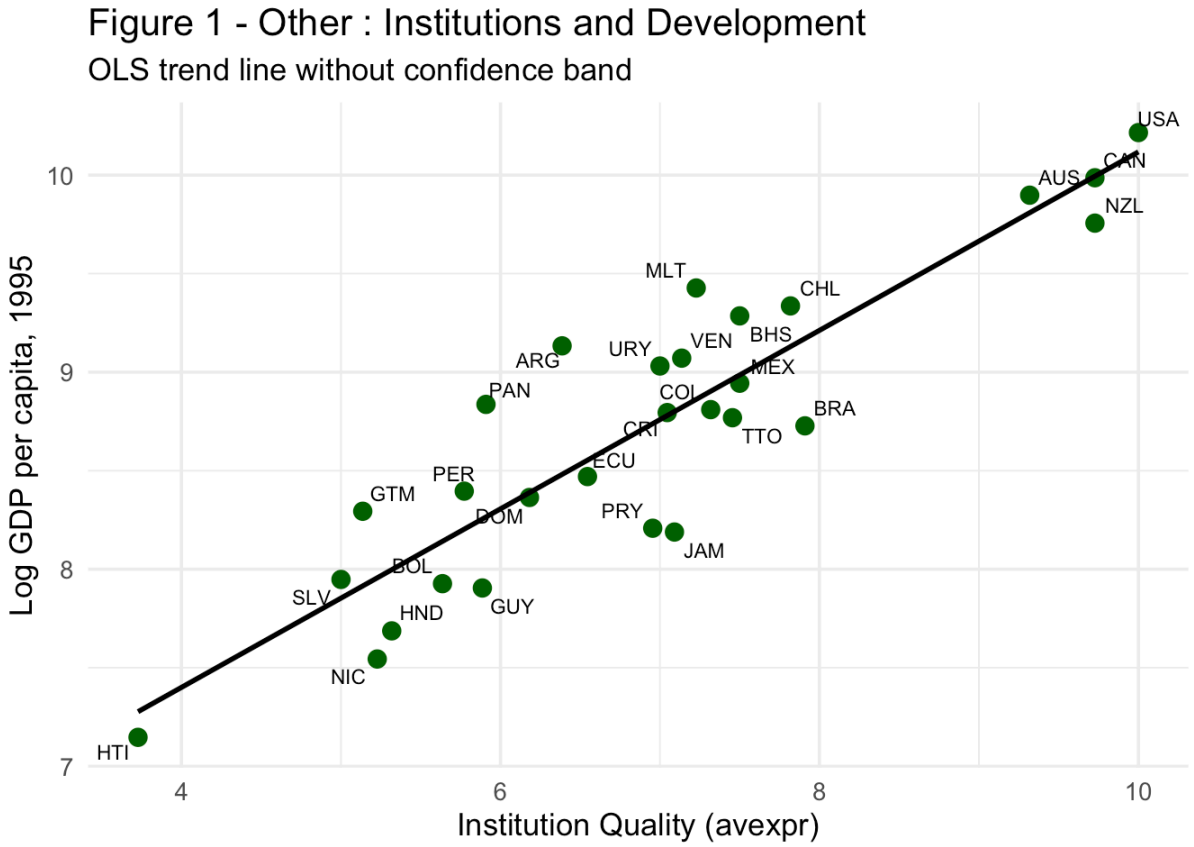
```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: label.
## ℹ This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



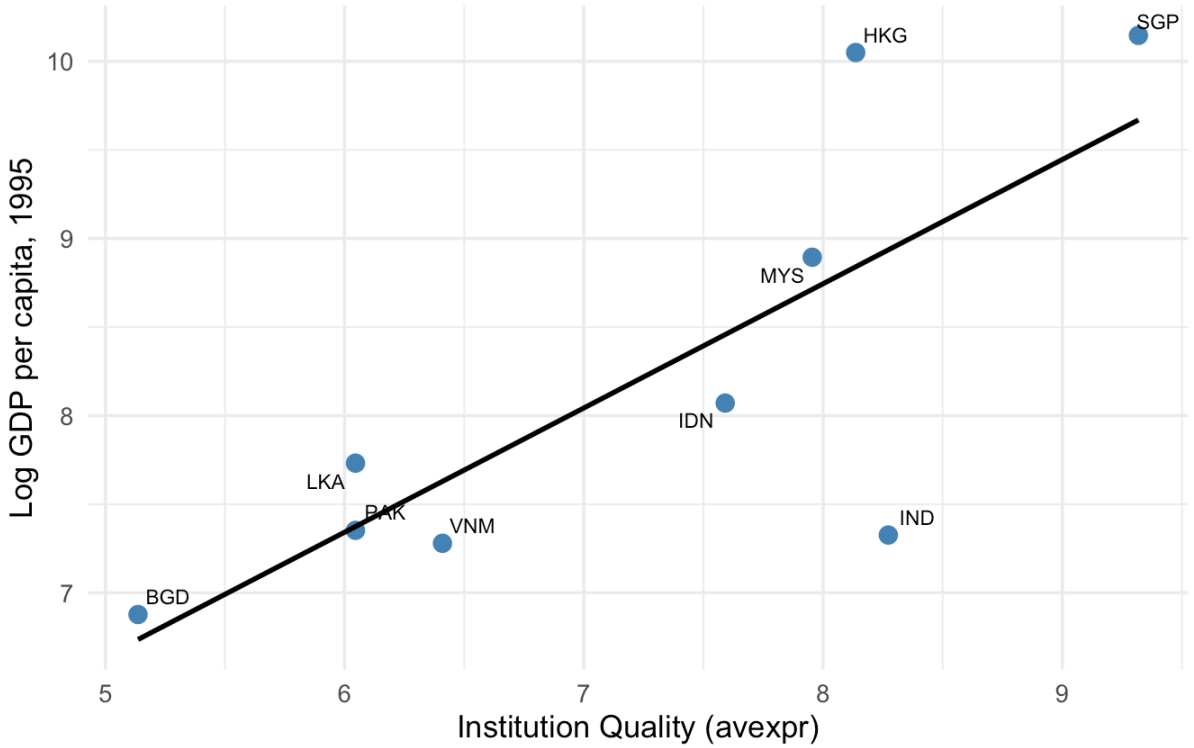

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: label.
## ℹ This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```




```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: label.
## ℹ This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```




```
## part 5

# --- Load and Prepare Data ---
ajr_dta <- read_dta("~/Desktop/maketable3.dta")

# Create the main sample used in most regressions for Table 3
main_sample <- ajr_dta %>%
  filter(excolony == 1, !is.na(extmort4)) %>%
  mutate(euro1900 = euro1900 / 100)

# Create the smaller subsample for regressions that also require log GDP data
lpgp_sample <- main_sample %>%
  filter(!is.na(logpgp95))

# --- Run Regressions for Panel A ---
panel_A_models <- list(
  "(1)" = feols(avexpr ~ cons00a, data = main_sample),
  "(2)" = feols(avexpr ~ cons00a + lat_abst, data = main_sample),
  "(3)" = feols(avexpr ~ democ00a, data = main_sample),
  "(4)" = feols(avexpr ~ democ00a + lat_abst, data = main_sample),
  "(5)" = feols(avexpr ~ indtime + cons1, data = main_sample),
  "(6)" = feols(avexpr ~ indtime + cons1 + lat_abst, data = main_sample),
  "(7)" = feols(avexpr ~ euro1900, data = main_sample),
  "(8)" = feols(avexpr ~ euro1900 + lat_abst, data = main_sample),
  "(9)" = feols(avexpr ~ logem4, data = lpgp_sample), # Uses smaller sample
  "(10)" = feols(avexpr ~ logem4 + lat_abst, data = lpgp_sample) # Uses smaller sample
)
```

```
## NOTE: 22 observations removed because of NA values (LHS: 18, RHS: 10).
## NOTE: 22 observations removed because of NA values (LHS: 18, RHS: 10).
```

```
## NOTE: 23 observations removed because of NA values (LHS: 18, RHS: 13).
## NOTE: 23 observations removed because of NA values (LHS: 18, RHS: 13).
```

```
## NOTE: 22 observations removed because of NA values (LHS: 18, RHS: 12).
## NOTE: 22 observations removed because of NA values (LHS: 18, RHS: 12).
```

```
## NOTE: 19 observations removed because of NA values (LHS: 18, RHS: 2).
## NOTE: 19 observations removed because of NA values (LHS: 18, RHS: 2).
```

```
## NOTE: 16 observations removed because of NA values (LHS: 16, RHS: 5).
## NOTE: 16 observations removed because of NA values (LHS: 16, RHS: 5).
```

```
# --- Generate Table for Panel A ---
modelsummary(
  panel_A_models,
  output = "gt",
  title = "Table 3, Panel A: Determinants of Institutions",
  coef_map = c("cons00a" = "Constraint on Executive in 1900", "democ00a" = "Democracy in 1900",
               "cons1" = "Constraint on Executive at Independence", "indtime" = "Date of Independe
nce",
               "euro1900" = "European Settlements in 1900", "logem4" = "Log Settler Mortality",
               "lat_abst" = "Distance from Equator"),
  gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
                 list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
  stars = TRUE,
  notes = "Notes: Dependent Variable is Average Expropriation Risk, 1985-95. Standard errors in pa
rentheses."
)
```

Table 3, Panel A: Determinants of Institutions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10
Constraint on Executive in 1900	0.321***	0.262**								
	(0.081)	(0.089)								
Democracy in 1900			0.244***	0.209**						
			(0.057)	(0.066)						
Constraint on Executive at Independence					0.247**	0.220**				
					(0.076)	(0.075)				
Date of Independence					0.009**	0.006+				

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Dependent Variable is Average Expropriation Risk, 1985-95. Standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10
					(0.003)	(0.003)				
European Settlements in 1900							3.177***	2.996***		
							(0.612)	(0.776)		
Log Settler Mortality									-0.607***	-0.51(
									(0.127)	(0.14
Distance from Equator		2.179		1.552		2.718+		0.580		2.0(
		(1.416)		(1.471)		(1.429)		(1.514)		(1.33
Num. Obs.	63	63	62	62	63	63	66	66	64	64
R-squared	0.203	0.233	0.236	0.251	0.194	0.241	0.296	0.298	0.270	0.29

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# --- Part 5 Run Regressions for Panel B ---
panel_B_models <- list(
  # Dep Var: Constraint on Executive in 1900
  "(1)" = feols(cons00a ~ euro1900, data = lpgp_sample),
  "(2)" = feols(cons00a ~ euro1900 + lat_abst, data = lpgp_sample),
  "(3)" = feols(cons00a ~ logem4, data = main_sample),
  "(4)" = feols(cons00a ~ logem4 + lat_abst, data = main_sample),
  # Dep Var: Democracy in 1900
  "(5)" = feols(democ00a ~ euro1900, data = lpgp_sample),
  "(6)" = feols(democ00a ~ euro1900 + lat_abst, data = lpgp_sample),
  "(7)" = feols(democ00a ~ logem4, data = lpgp_sample),
  "(8)" = feols(democ00a ~ logem4 + lat_abst, data = lpgp_sample),
  # Dep Var: European Settlements in 1900
  "(9)" = feols(euro1900 ~ logem4, data = lpgp_sample),
  "(10)" = feols(euro1900 ~ logem4 + lat_abst, data = lpgp_sample)
)
```

```
## NOTE: 10 observations removed because of NA values (LHS: 9, RHS: 2).
## NOTE: 10 observations removed because of NA values (LHS: 9, RHS: 2).
```

```
## NOTE: 10 observations removed because of NA values (LHS: 10, RHS: 5).
## NOTE: 10 observations removed because of NA values (LHS: 10, RHS: 5).
```

```
## NOTE: 13 observations removed because of NA values (LHS: 12, RHS: 2).
## NOTE: 13 observations removed because of NA values (LHS: 12, RHS: 2).
```

```
## NOTE: 12 observations removed because of NA values (LHS: 12, RHS: 5).
## NOTE: 12 observations removed because of NA values (LHS: 12, RHS: 5).
```

```
## NOTE: 7 observations removed because of NA values (LHS: 2, RHS: 5).
## NOTE: 7 observations removed because of NA values (LHS: 2, RHS: 5).
```

```
# --- Generate Table for Panel B ---
modelsummary(
  panel_B_models,
  output = "gt",
  title = "Table 3, Panel B: Determinants of Early Institutions",
  coef_map = c("euro1900" = "European Settlements in 1900", "logem4" = "Log Settler Mortality",
               "lat_abst" = "Distance from Equator"),
  gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
                 list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
  stars = TRUE,
  notes = "Notes: Standard errors in parentheses."
) %>%
  # Add spanners to clarify the dependent variable for each set of columns
  tab_spanner(label = "Dep. Var: Constraint on Executive in 1900", columns = 2:5) %>%
  tab_spanner(label = "Dep. Var: Democracy in 1900", columns = 6:9) %>%
  tab_spanner(label = "Dep. Var: European Settlements in 1900", columns = 10:11)
```

Table 3, Panel B: Determinants of Early Institutions

	Dep. Var: Constraint on Executive in 1900				Dep. Var: Democracy in 1900				Dep. Var: E Settlements	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
European Settlements in 1900	5.491***	5.385***			8.566***	8.056***				
	(0.727)	(0.929)			(0.932)	(1.190)				
Log Settler Mortality			-0.821***	-0.654***			-1.221***	-0.877***	-0.112***	-
			(0.167)	(0.181)			(0.242)	(0.253)	(0.020)	
Distance from Equator		0.333		3.634*		1.605		7.569**		(
		(1.809)		(1.716)		(2.315)		(2.425)		
Num. Obs.	70	70	75	75	67	67	68	68	73	
R-squared	0.456	0.456	0.248	0.292	0.565	0.568	0.278	0.372	0.305	

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Standard errors in parentheses.

```
## Part 6

# --- Load Data ---
ajr_dta <- read_dta("~/Desktop/maketable4.dta") %>%
  as.data.frame()   # fixest prefer data.frame

# --- Create Base Sample ---
base_sample <- ajr_dta %>%
  filter(baseco == 1) %>%
  mutate(
    other_cont = if_else(shortnam %in% c("AUS", "MLT", "NZL"), 1, 0)
  )

# --- Create Subsamples ---
no_neo_europes_sample <- base_sample %>% filter(rich4 != 1)
no_africa_sample <- base_sample %>% filter(africa != 1)

# --- IV Regression Models ---
iv_models <- list(
  "(1)" = feols(logpgp95 ~ 1 | avexpr ~ logem4, data = base_sample),
  "(2)" = feols(logpgp95 ~ lat_abst | avexpr ~ logem4, data = base_sample),
  "(3)" = feols(logpgp95 ~ 1 | avexpr ~ logem4, data = no_neo_europes_sample),
  "(4)" = feols(logpgp95 ~ lat_abst | avexpr ~ logem4, data = no_neo_europes_sample),
  "(5)" = feols(logpgp95 ~ 1 | avexpr ~ logem4, data = no_africa_sample),
  "(6)" = feols(logpgp95 ~ lat_abst | avexpr ~ logem4, data = no_africa_sample),
  "(7)" = feols(logpgp95 ~ africa + asia + other_cont | avexpr ~ logem4, data = base_sample),
  "(8)" = feols(logpgp95 ~ lat_abst + africa + asia + other_cont | avexpr ~ logem4, data = base_sa
mple),
  "(9)" = feols(loghjypl ~ 1 | avexpr ~ logem4, data = base_sample)
)
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3).
```

```
# --- Generate Table for Panel A (2SLS Results) ---
modelsummary(
  iv_models,
  output = "gt",
  title = "Table 4, Panel A: Two-Stage Least Squares Estimates",
  coef_map = c("fit_avexpr" = "Average protection against expropriation risk 1985-1995",
               "lat_abst"   = "Latitude",
               "africa"     = "Africa dummy",
               "asia"       = "Asia dummy",
               "other_cont" = "'Other' continent dummy"),
  gof_map = "nobs",
  stars = TRUE,
  notes = "Notes: 2SLS estimates with standard errors in parentheses."
)
```

Table 4, Panel A: Two-Stage Least Squares Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: 2SLS estimates with standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average protection against expropriation risk 1985-1995	0.944***	0.996***	1.281***	1.212**	0.578***	0.576***	0.982**	1.107*	0.981***
	(0.157)	(0.222)	(0.358)	(0.354)	(0.098)	(0.117)	(0.299)	(0.464)	(0.171)
Latitude		-0.647		0.939		0.038		-1.178	
		(1.335)		(1.463)		(0.835)		(1.755)	
Africa dummy							-0.464	-0.437	
							(0.358)	(0.424)	
Asia dummy							-0.924*	-1.047+	
							(0.400)	(0.525)	
'Other' continent dummy							-0.941	-0.990	
							(0.848)	(0.998)	
Num.Obs.	64	64	60	60	37	37	64	64	61

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: 2SLS estimates with standard errors in parentheses.

```
# ---  Generate Table for Panel B (First Stage Results) ---

first_stage_models <- purrr::map(iv_models, ~.$iv_first_stage$avexpr)
# ----------------------------------

modelsummary(
  first_stage_models,
  output = "gt",
  title = "Table 4, Panel B: First Stage for Average Protection Against Expropriation Risk",
  coef_map = c("logem4"     = "Log European settler mortality",
               "lat_abst"   = "Latitude",
               "africa"     = "Africa dummy",
               "asia"       = "Asia dummy",
               "other_cont" = "'Other' continent dummy",
               "(Intercept)" = "Constant"),
  gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
                 list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
  stars = TRUE,
  notes = "Notes: First-stage OLS regressions. Dependent variable is Expropriation Risk. Standard
errors in parentheses."
)
```

Table 4, Panel B: First Stage for Average Protection Against Expropriation Risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: First-stage OLS regressions. Dependent variable is Expropriation Risk. Standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log European settler mortality	-0.607***	-0.510***	-0.391**	-0.394**	-1.210***	-1.137***	-0.432*	-0.340+	-0.629***
	(0.127)	(0.141)	(0.133)	(0.141)	(0.219)	(0.245)	(0.173)	(0.183)	(0.129)
Latitude		2.002		-0.106		0.990		2.009	
		(1.337)		(1.487)		(1.428)		(1.391)	
Africa dummy							-0.269	-0.258	
							(0.413)	(0.410)	
Asia dummy							0.333	0.472	
							(0.498)	(0.503)	
'Other' continent dummy							1.241	1.062	
							(0.842)	(0.844)	
Constant	9.341***	8.529***	8.184***	8.216***	11.844***	11.344***	8.538***	7.729***	9.456***
	(0.611)	(0.812)	(0.657)	(0.800)	(0.898)	(1.157)	(0.783)	(0.957)	(0.625)
Num. Obs.	64	64	60	60	37	37	64	64	61
R-squared	0.270	0.296	0.130	0.130	0.466	0.473	0.304	0.328	0.287

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: First-stage OLS regressions. Dependent variable is Expropriation Risk. Standard errors in parentheses.

```
# --- Run Regressions for Panel C ---
panel_C_models <- list(
  "(1)" = feols(logpgp95 ~ avexpr, data = base_sample),
  "(2)" = feols(logpgp95 ~ avexpr + lat_abst, data = base_sample),
  "(3)" = feols(logpgp95 ~ avexpr, data = no_neo_europes_sample),
  "(4)" = feols(logpgp95 ~ avexpr + lat_abst, data = no_neo_europes_sample),
  "(5)" = feols(logpgp95 ~ avexpr, data = no_africa_sample),
  "(6)" = feols(logpgp95 ~ avexpr + lat_abst, data = no_africa_sample),
  "(7)" = feols(logpgp95 ~ avexpr + africa + asia + other_cont, data = base_sample),
  "(8)" = feols(logpgp95 ~ avexpr + lat_abst + africa + asia + other_cont, data = base_sample),
  "(9)" = feols(loghjypl ~ avexpr, data = base_sample)
)
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3).
```

```
# --- Generate Table for Panel C ---
modelsummary(
  panel_C_models,
  output = "gt",
  title = "Table 4, Panel C: OLS Regressions",
  coef_map = c("avexpr" = "Protection Against Expropriation Risk", "lat_abst" = "Distance from Equ
ator",
               "africa" = "Africa", "asia" = "Asia", "other_cont" = "Other Continents"),
  gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
                 list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
  stars = TRUE,
  notes = "Notes: OLS regressions with standard errors in parentheses."
)
```

Table 4, Panel C: OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Protection Against Expropriation Risk	0.522***	0.468***	0.487***	0.471***	0.482***	0.466***	0.424***	0.401***	0.457***
	(0.061)	(0.064)	(0.076)	(0.074)	(0.065)	(0.071)	(0.057)	(0.059)	(0.061)
Distance from Equator		1.577*		1.802*		0.462		0.875	
		(0.710)		(0.852)		(0.731)		(0.628)	
Africa							-0.918***	-0.881***	
							(0.169)	(0.170)	
Asia							-0.631**	-0.577*	
							(0.230)	(0.231)	
Other Continents							0.216	0.107	
							(0.377)	(0.382)	
Num. Obs.	64	64	60	60	37	37	64	64	61
R-squared	0.540	0.574	0.414	0.456	0.611	0.616	0.704	0.714	0.486

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: OLS regressions with standard errors in parentheses.

}

```
# ---  Part 6 Generate Table for Panel B (First Stage Results) ---

first_stage_models <- purrr::map(iv_models, ~.$iv_first_stage$avexpr)
# ---------------------------------

modelsummary(
  first_stage_models,
  output = "gt",
  title = "Table 4, Panel B: First Stage for Average Protection Against Expropriation Risk",
  coef_map = c("logem4"     = "Log European settler mortality",
               "lat_abst"   = "Latitude",
               "africa"     = "Africa dummy",
               "asia"       = "Asia dummy",
               "other_cont" = "'Other' continent dummy",
               "(Intercept)" = "Constant"),
  gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
                 list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
  stars = TRUE,
  notes = "Notes: First-stage OLS regressions. Dependent variable is Expropriation Risk. Standard
errors in parentheses."
)
```

Table 4, Panel B: First Stage for Average Protection Against Expropriation Risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log European settler mortality	-0.607***	-0.510***	-0.391**	-0.394**	-1.210***	-1.137***	-0.432*	-0.340+	-0.629***
	(0.127)	(0.141)	(0.133)	(0.141)	(0.219)	(0.245)	(0.173)	(0.183)	(0.129)
Latitude		2.002		-0.106		0.990		2.009	
		(1.337)		(1.487)		(1.428)		(1.391)	
Africa dummy							-0.269	-0.258	
							(0.413)	(0.410)	
Asia dummy							0.333	0.472	
							(0.498)	(0.503)	
'Other' continent dummy							1.241	1.062	
							(0.842)	(0.844)	
Constant	9.341***	8.529***	8.184***	8.216***	11.844***	11.344***	8.538***	7.729***	9.456***
	(0.611)	(0.812)	(0.657)	(0.800)	(0.898)	(1.157)	(0.783)	(0.957)	(0.625)

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: First-stage OLS regressions. Dependent variable is Expropriation Risk. Standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Num. Obs.	64	64	60	60	37	37	64	64	61
R-squared	0.270	0.296	0.130	0.130	0.466	0.473	0.304	0.328	0.287

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: First-stage OLS regressions. Dependent variable is Expropriation Risk. Standard errors in parentheses.

```
# --- Run Regressions for Panel C ---
panel_C_models <- list(
  "(1)" = feols(logpgp95 ~ avexpr, data = base_sample),
  "(2)" = feols(logpgp95 ~ avexpr + lat_abst, data = base_sample),
  "(3)" = feols(logpgp95 ~ avexpr, data = no_neo_europes_sample),
  "(4)" = feols(logpgp95 ~ avexpr + lat_abst, data = no_neo_europes_sample),
  "(5)" = feols(logpgp95 ~ avexpr, data = no_africa_sample),
  "(6)" = feols(logpgp95 ~ avexpr + lat_abst, data = no_africa_sample),
  "(7)" = feols(logpgp95 ~ avexpr + africa + asia + other_cont, data = base_sample),
  "(8)" = feols(logpgp95 ~ avexpr + lat_abst + africa + asia + other_cont, data = base_sample),
  "(9)" = feols(loghjypl ~ avexpr, data = base_sample)
)
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3).
```

```
# --- Generate Table for Panel C ---
modelsummary(
  panel_C_models,
  output = "gt",
  title = "Table 4, Panel C: OLS Regressions",
  coef_map = c("avexpr" = "Protection Against Expropriation Risk", "lat_abst" = "Distance from Equ
ator",
               "africa" = "Africa", "asia" = "Asia", "other_cont" = "Other Continents"),
  gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
                 list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
  stars = TRUE,
  notes = "Notes: OLS regressions with standard errors in parentheses."
)
```

Table 4, Panel C: OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Protection Against Expropriation Risk	0.522***	0.468***	0.487***	0.471***	0.482***	0.466***	0.424***	0.401***	0.457***
	(0.061)	(0.064)	(0.076)	(0.074)	(0.065)	(0.071)	(0.057)	(0.059)	(0.061)
Distance from Equator		1.577*		1.802*		0.462		0.875	
		(0.710)		(0.852)		(0.731)		(0.628)	

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: OLS regressions with standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Africa							-0.918***	-0.881***	
							(0.169)	(0.170)	
Asia							-0.631**	-0.577*	
							(0.230)	(0.231)	
Other Continents							0.216	0.107	
							(0.377)	(0.382)	
Num. Obs.	64	64	60	60	37	37	64	64	61
R-squared	0.540	0.574	0.414	0.456	0.611	0.616	0.704	0.714	0.486

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: OLS regressions with standard errors in parentheses.

```
# ---  Load and Prepare Data ---
ajr_dta <- read_dta("~/Desktop/maketable5.dta")

# Create the base sample
base_sample <- ajr_dta %>% filter(baseco == 1)

# Create the British colonies subsample
brit_colonies_sample <- base_sample %>% filter(f_brit == 1)
```

```
# --- Run IV and OLS Regressions (Part 7) ---

# IV models (Panels A & B)

iv_models_p5 <- list(
"(1)" = feols(logpgp95 ~ f_brit + f_french | avexpr ~ logem4, data = base_sample),
"(2)" = feols(logpgp95 ~ lat_abst + f_brit + f_french | avexpr ~ logem4, data = base_sample),
"(3)" = feols(logpgp95 ~ 1 | avexpr ~ logem4, data = brit_colonies_sample),
"(4)" = feols(logpgp95 ~ lat_abst | avexpr ~ logem4, data = brit_colonies_sample),
"(5)" = feols(logpgp95 ~ sjlofr | avexpr ~ logem4, data = base_sample),
"(6)" = feols(logpgp95 ~ lat_abst + sjlofr | avexpr ~ logem4, data = base_sample),
"(7)" = feols(logpgp95 ~ catho80 + muslim80 + no_cpm80 | avexpr ~ logem4, data = base_sample),
"(8)" = feols(logpgp95 ~ lat_abst + catho80 + muslim80 + no_cpm80 | avexpr ~ logem4, data = base_s
ample),
"(9)" = feols(logpgp95 ~ f_french + sjlofr + catho80 + muslim80 + no_cpm80 | avexpr ~ logem4, data
= base_sample),
"(10)" = feols(logpgp95 ~ lat_abst + f_french + sjlofr + catho80 + muslim80 + no_cpm80 | avexpr ~
logem4, data = base_sample)
)

# OLS models (Panel C)

ols_models_p5 <- list(
"(1)" = feols(logpgp95 ~ avexpr + f_brit + f_french, data = base_sample),
"(2)" = feols(logpgp95 ~ avexpr + lat_abst + f_brit + f_french, data = base_sample),
"(3)" = feols(logpgp95 ~ avexpr, data = brit_colonies_sample),
"(4)" = feols(logpgp95 ~ avexpr + lat_abst, data = brit_colonies_sample),
"(5)" = feols(logpgp95 ~ avexpr + sjlofr, data = base_sample),
"(6)" = feols(logpgp95 ~ avexpr + lat_abst + sjlofr, data = base_sample),
"(7)" = feols(logpgp95 ~ avexpr + catho80 + muslim80 + no_cpm80, data = base_sample),
"(8)" = feols(logpgp95 ~ avexpr + lat_abst + catho80 + muslim80 + no_cpm80, data = base_sample),
"(9)" = feols(logpgp95 ~ avexpr + lat_abst + f_french + sjlofr + catho80 + muslim80 + no_cpm80, da
ta = base_sample)
)
```

```
# Panel A: IV regressions (Table 5, Panel A)

modelsummary(
iv_models_p5,
output = "gt",
title = "Table 5, Panel A: IV Regressions of Log GDP Per Capita with Additional Controls",
coef_map = c("fit_avexpr" = "Average Expropriation Risk", "lat_abst" = "Latitude",
"f_brit" = "British Colony Dummy", "f_french" = "French Colony Dummy",
"sjlofr" = "French Legal Origin", "catho80" = "Catholic Religion Dummy",
"muslim80" = "Muslim Religion Dummy", "no_cpm80" = "Other Religion Dummy"),
gof_map = "nobs",
stars = TRUE
)
```

Table 5, Panel A: IV Regressions of Log GDP Per Capita with Additional Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001										

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average Expropriation Risk	1.078***	1.155**	1.066***	1.339*	1.080***	1.181***	0.917***	1.006***	1.012***	1.212**
	(0.218)	(0.337)	(0.244)	(0.516)	(0.191)	(0.291)	(0.147)	(0.252)	(0.187)	(0.395)
Latitude		-0.751		-2.986		-1.126		-0.938		-1.794
		(1.675)		(3.214)		(1.560)		(1.503)		(2.133)
British Colony Dummy	-0.778*	-0.795*								
	(0.354)	(0.393)								
French Colony Dummy	-0.117	-0.058							0.290	0.396
	(0.355)	(0.419)							(0.378)	(0.495)
French Legal Origin					0.887**	0.962*			0.234	0.294
					(0.324)	(0.394)			(0.411)	(0.519)
Catholic Religion Dummy							0.009	0.008	0.007	0.005
							(0.009)	(0.010)	(0.011)	(0.014)
Muslim Religion Dummy							0.000	0.000	-0.002	-0.003
							(0.009)	(0.010)	(0.010)	(0.012)
Other Religion Dummy							-0.007	-0.010	-0.008	-0.014
							(0.010)	(0.012)	(0.011)	(0.015)
Num.Obs.	64	64	25	25	64	64	64	64	64	64

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# Panel B: First stage

first_stage_models_p5 <- purrr::map(iv_models_p5, ~.$iv_first_stage$avexpr)

modelsummary(
first_stage_models_p5,
output = "gt",
title = "Table 5, Panel B: First-Stage Regressions",
coef_map = c("logem4" = "Log Settler Mortality", "lat_abst" = "Latitude",
"f_brit" = "British Colony Dummy", "f_french" = "French Colony Dummy",
"sjlofr" = "French Legal Origin", "catho80" = "Catholic Religion Dummy",
"muslim80" = "Muslim Religion Dummy", "no_cpm80" = "Other Religion Dummy",
"(Intercept)" = "Constant"),
gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
stars = TRUE
)
```

Table 5, Panel B: First-Stage Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(
Log Settler Mortality	-0.534***	-0.426**	-0.593**	-0.419+	-0.544***	-0.443**	-0.579***	-0.442**	-0.544***	-0
	(0.140)	(0.158)	(0.189)	(0.213)	(0.127)	(0.141)	(0.130)	(0.151)	(0.141)	(0
Latitude		1.970		3.202		2.079		2.496+		2.
		(1.375)		(2.002)		(1.301)		(1.453)		(1.
British Colony Dummy	0.629+	0.548								
	(0.366)	(0.368)								
French Colony Dummy	0.047	-0.117							-0.004	-0
	(0.430)	(0.441)							(0.524)	(0
French Legal Origin					-0.674*	-0.688*			-0.408	-0
					(0.328)	(0.324)			(0.552)	(0.
Catholic Religion Dummy							-0.018	-0.012	-0.012	-0
							(0.013)	(0.013)	(0.016)	(0

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(
Muslim Religion Dummy							-0.016	-0.012	-0.011	-0
							(0.012)	(0.012)	(0.013)	(0
Other Religion Dummy							-0.009	0.000	-0.006	0.
							(0.015)	(0.016)	(0.016)	(0
Constant	8.747***	7.958***	9.617***	8.216***	9.471***	8.631***	10.557***	8.877***	10.196***	8.3
	(0.690)	(0.878)	(0.835)	(1.192)	(0.599)	(0.791)	(1.292)	(1.604)	(1.363)	(1.
Num. Obs.	64	64	25	25	64	64	64	64	64	
R-squared	0.308	0.331	0.300	0.373	0.317	0.345	0.321	0.354	0.331	0.

```
# Panel C: OLS

modelsummary(
ols_models_p5,
output = "gt",
title = "Table 5, Panel C: OLS Regressions",
coef_map = c("avexpr" = "Average Expropriation Risk"),
gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
stars = TRUE
)
```

Table 5, Panel C: OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average Expropriation Risk	0.528***	0.465***	0.606***	0.549***	0.563***	0.508***	0.533***	0.471***	0.471***
	(0.065)	(0.067)	(0.093)	(0.108)	(0.064)	(0.067)	(0.054)	(0.057)	(0.060)
Num. Obs.	64	64	25	25	64	64	64	64	64
R-squared	0.565	0.610	0.648	0.664	0.566	0.597	0.691	0.722	0.722

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# --- Load and Prepare Data for Part 8 ---

ajr_dta6 <- read_dta("~/Desktop/maketable6.dta")
base_sample6 <- ajr_dta6 %>% filter(baseco == 1)
```

```
# --- Define control sets ---

temp_humid_controls <- "temp1 + temp2 + humid1 + humid2"
resource_controls <- "steplow + deslow + stepmid + desmid + drystep + drywint + goldm + iron + sil
v + zinc + oilres + landlock"
all_controls <- paste("lat_abst", temp_humid_controls, "edes1975", "avelf", resource_controls, sep
= " + ")
```

```
# IV models for Part 8 (Table 6)

iv_models_p6 <- list(
"(1)" = feols(logpgp95 ~ 1 | avexpr ~ logem4, data = base_sample6),
"(2)" = feols(as.formula(paste("logpgp95 ~ lat_abst +", temp_humid_controls, "| avexpr ~ logem
4")), data = base_sample6),
"(3)" = feols(logpgp95 ~ edes1975 | avexpr ~ logem4, data = base_sample6),
"(4)" = feols(logpgp95 ~ lat_abst + edes1975 | avexpr ~ logem4, data = base_sample6),
"(5)" = feols(as.formula(paste("logpgp95 ~", resource_controls, "| avexpr ~ logem4")), data = base
_sample6),
"(6)" = feols(as.formula(paste("logpgp95 ~ lat_abst +", resource_controls, "| avexpr ~ logem4")),
data = base_sample6),
"(7)" = feols(logpgp95 ~ avelf | avexpr ~ logem4, data = base_sample6),
"(8)" = feols(logpgp95 ~ lat_abst + avelf | avexpr ~ logem4, data = base_sample6),
"(9)" = feols(as.formula(paste("logpgp95 ~", all_controls, "| avexpr ~ logem4")), data = base_samp
le6)
)

# OLS models for Part 8 (Panel C)

ols_models_p6 <- list(
"(1)" = feols(as.formula(paste("logpgp95 ~ avexpr +", temp_humid_controls)), data = base_sample6),
"(2)" = feols(as.formula(paste("logpgp95 ~ avexpr + lat_abst +", temp_humid_controls)), data = bas
e_sample6),
"(3)" = feols(logpgp95 ~ avexpr + edes1975, data = base_sample6),
"(4)" = feols(logpgp95 ~ avexpr + lat_abst + edes1975, data = base_sample6),
"(5)" = feols(as.formula(paste("logpgp95 ~ avexpr +", resource_controls)), data = base_sample6),
"(6)" = feols(as.formula(paste("logpgp95 ~ avexpr + lat_abst +", resource_controls)), data = base_
sample6),
"(7)" = feols(logpgp95 ~ avexpr + avelf, data = base_sample6),
"(8)" = feols(logpgp95 ~ avexpr + lat_abst + avelf, data = base_sample6),
"(9)" = feols(as.formula(paste("logpgp95 ~ avexpr +", all_controls)), data = base_sample6)
)
```

```
# Panel A: Two-stage least squares (Table 6, Panel A)

# Build nicely formatted custom table using gt if desired.

# Here we use a custom, manual build (as in original script) and also provide standard modelsummar
y as fallback.

# Standard modelsummary output:

modelsummary(
iv_models_p6,
output = "gt",
title = "Table 6, Panel A: Two-Stage Least Squares (standard layout)",
coef_map = c("fit_avexpr" = "Average Expropriation Risk"),
gof_map = "nobs",
stars = TRUE
)
```

```
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
```

Table 6, Panel A: Two-Stage Least Squares (standard layout)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average Expropriation Risk	0.944***	0.997***	0.960***	0.991**	1.259**	1.358*	0.738***	0.787***	0.781**
	(0.157)	(0.246)	(0.277)	(0.300)	(0.439)	(0.618)	(0.130)	(0.165)	(0.225)
Num.Obs.	64	64	64	64	64	64	64	64	64

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# Panel B: First-stage regressions (Table 6, Panel B)

first_stage_p6 <- purrr::map(iv_models_p6, ~.$iv_first_stage$avexpr)

modelsummary(
first_stage_p6,
output = "gt",
title = "Table 6, Panel B: First-Stage Regressions",
coef_map = c("logem4" = "Log Settler Mortality"),
gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
stars = TRUE,
notes = "Notes: Only the coefficient on Log Settler Mortality is shown. All models include the ful
l set of controls."
)
```

```
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
```

Table 6, Panel B: First-Stage Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log Settler Mortality	-0.607***	-0.546**	-0.413**	-0.404**	-0.368*	-0.304+	-0.636***	-0.556***	-0.541**
	(0.127)	(0.158)	(0.143)	(0.147)	(0.157)	(0.167)	(0.146)	(0.152)	(0.198)
Num. Obs.	64	64	64	64	64	64	64	64	64
R-squared	0.270	0.381	0.342	0.343	0.455	0.468	0.272	0.304	0.566

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Only the coefficient on Log Settler Mortality is shown. All models include the full set of controls.

```r
# Panel C: OLS robustness checks

modelsummary(
ols_models_p6,
output = "gt",
title = "Table 6, Panel C: OLS Robustness Checks",
coef_map = c("avexpr" = "Average Expropriation Risk"),
gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
stars = TRUE,
notes = "Notes: Only the coefficient on Average Expropriation Risk is shown. All models include th
e full set of controls."
)
```

```
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
## Warning: The VCOV matrix is not positive semi-definite and was 'fixed' (see
## ?vcov).
```

Table 6, Panel C: OLS Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average Expropriation Risk	0.454***	0.406***	0.386***	0.384***	0.446***	0.410***	0.462***	0.450***	0.398***
	(0.065)	(0.067)	(0.062)	(0.063)	(0.077)	(0.077)	(0.052)	(0.055)	(0.064)
Num. Obs.	64	64	64	64	64	64	64	64	64
R-squared	0.593	0.625	0.650	0.650	0.645	0.669	0.687	0.689	0.823

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Only the coefficient on Average Expropriation Risk is shown. All models include the full set of controls.

```
library(haven)
library(dplyr)


ajr_dta7 <- read_dta("~/Desktop/maketable7.dta")


base_sample7 <- ajr_dta7 %>%
  filter(baseco == 1) %>%
  mutate(other_cont = if_else(shortnam %in% c('colonial', 'non-colonial'), 1, 0))


ols_sample_789 <- base_sample7 %>%
  filter(!is.na(logem4), !is.na(latabs) | !is.na(lat_abst), !is.na(lt100km), !is.na(meantemp))

ols_sample_1011 <- base_sample7 %>%
  filter(!is.na(yellow))
```

```
# IV models (Table 7)

iv_models_p7 <- list(
"(1)" = feols(logpgp95 ~ malfal94 | avexpr ~ logem4, data = base_sample7),
"(2)" = feols(logpgp95 ~ lat_abst + malfal94 | avexpr ~ logem4, data = base_sample7),
"(3)" = feols(logpgp95 ~ leb95 | avexpr ~ logem4, data = base_sample7),
"(4)" = feols(logpgp95 ~ lat_abst + leb95 | avexpr ~ logem4, data = base_sample7),
"(5)" = feols(logpgp95 ~ imr95 | avexpr ~ logem4, data = base_sample7),
"(6)" = feols(logpgp95 ~ lat_abst + imr95 | avexpr ~ logem4, data = base_sample7),
"(7)" = feols(logpgp95 ~ 1 | avexpr + malfal94 ~ logem4 + latabs + lt100km + meantemp, data = base
_sample7),
"(8)" = feols(logpgp95 ~ 1 | avexpr + leb95 ~ logem4 + latabs + lt100km + meantemp, data = base_sa
mple7),
"(9)" = feols(logpgp95 ~ 1 | avexpr + imr95 ~ logem4 + latabs + lt100km + meantemp, data = base_sa
mple7),
"(10)" = feols(logpgp95 ~ 1 | avexpr ~ yellow, data = base_sample7),
"(11)" = feols(logpgp95 ~ africa + asia + other_cont | avexpr ~ yellow, data = base_sample7)
)
```

```
## NOTE: 2 observations removed because of NA values (RHS: 2).
## NOTE: 2 observations removed because of NA values (RHS: 2).
```

```
## NOTE: 4 observations removed because of NA values (RHS: 4).
## NOTE: 4 observations removed because of NA values (RHS: 4).
## NOTE: 4 observations removed because of NA values (RHS: 4).
## NOTE: 4 observations removed because of NA values (RHS: 4).
```

```
## NOTE: 4 observations removed because of NA values (IV: 2/4).
```

```
## NOTE: 5 observations removed because of NA values (IV: 4/4).
## NOTE: 5 observations removed because of NA values (IV: 4/4).
```

```
## The exogenous variable 'other_cont' have been removed because of collinearity
## (see $collin.var).
```

```
# OLS models (Panel C)

ols_models_p7 <- list(
"(1)" = feols(logpgp95 ~ avexpr + malfal94, data = base_sample7),
"(2)" = feols(logpgp95 ~ avexpr + lat_abst + malfal94, data = base_sample7),
"(3)" = feols(logpgp95 ~ avexpr + leb95, data = base_sample7),
"(4)" = feols(logpgp95 ~ avexpr + lat_abst + leb95, data = base_sample7),
"(5)" = feols(logpgp95 ~ avexpr + imr95, data = base_sample7),
"(6)" = feols(logpgp95 ~ avexpr + lat_abst + imr95, data = base_sample7),
"(7)" = feols(logpgp95 ~ avexpr + malfal94, data = ols_sample_789),
"(8)" = feols(logpgp95 ~ avexpr + leb95, data = ols_sample_789),
"(9)" = feols(logpgp95 ~ avexpr + imr95, data = ols_sample_789),
"(10)" = feols(logpgp95 ~ avexpr, data = ols_sample_1011),
"(11)" = feols(logpgp95 ~ avexpr + africa + asia + other_cont, data = ols_sample_1011)
)
```

```
## NOTE: 2 observations removed because of NA values (RHS: 2).
```

```
## NOTE: 2 observations removed because of NA values (RHS: 2).
```

```
## NOTE: 4 observations removed because of NA values (RHS: 4).
## NOTE: 4 observations removed because of NA values (RHS: 4).
## NOTE: 4 observations removed because of NA values (RHS: 4).
## NOTE: 4 observations removed because of NA values (RHS: 4).
```

```
## NOTE: 1 observation removed because of NA values (RHS: 1).
## NOTE: 1 observation removed because of NA values (RHS: 1).
```

```
## The variable 'other_cont' has been removed because of collinearity (see
## $collin.var).
```

```
# Panel A (Table 7, Panel A): 2SLS with geography and health variables

modelsummary(
iv_models_p7,
output = "gt",
title = "Table 7, Panel A: 2SLS Regressions with Geography and Health Variables",
coef_map = c("fit_avexpr" = "Average Expropriation Risk", "fit_malfal94" = "Malaria Index",
"fit_leb95" = "Life Expectancy", "fit_imr95" = "Infant Mortality",
"lat_abst" = "Latitude", "malfal94" = "Malaria Index",
"leb95" = "Life Expectancy", "imr95" = "Infant Mortality",
"africa" = "Africa Dummy", "asia" = "Asia Dummy", "other_cont" = "Other Continent Dummy"),
gof_map = "nobs",
stars = TRUE
)
```

Table 7, Panel A: 2SLS Regressions with Geography and Health Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11
Average Expropriation Risk	0.687**	0.721*	0.629*	0.677+	0.551*	0.562+	0.689*	0.737**	0.675**	0.914***	0.82
	(0.251)	(0.297)	(0.276)	(0.338)	(0.239)	(0.314)	(0.258)	(0.236)	(0.227)	(0.245)	(0.2
Malaria Index	-0.578	-0.599					-0.623				
	(0.467)	(0.473)					(0.685)				
Life Expectancy			0.027	0.026				0.017			
			(0.021)	(0.023)				(0.023)			
Infant Mortality					-0.010+	-0.010			-0.007		
					(0.005)	(0.006)			(0.007)		
Latitude		-0.565		-0.529		-0.102					
		(1.043)		(0.968)		(0.949)					
Africa Dummy											-0.5
											(0.3
Asia Dummy											-0.7
											(0.3
Num.Obs.	62	62	60	60	60	60	60	59	59	64	6

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# Panel B: first-stage for avexpr

first_stage_models_avexpr <- purrr::map(iv_models_p7, ~.$iv_first_stage$avexpr)

modelsummary(
first_stage_models_avexpr,
output = "gt",
title = "Table 7, Panel B: First-Stage Regressions for Expropriation Risk",
coef_map = c("logem4" = "Log Settler Mortality", "yellow" = "Yellow Fever Dummy",
"latabs" = "Latitude (abs)", "lt100km" = "Coastal Dummy",
"meantemp" = "Mean Temperature", "lat_abst" = "Latitude",
"malfal94" = "Malaria Index", "leb95" = "Life Expectancy", "imr95" = "Infant Mortality",
"africa" = "Africa Dummy", "asia" = "Asia Dummy", "other_cont" = "Other Continent Dummy",
"(Intercept)" = "Constant"),
gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
stars = TRUE
)
```

Table 7, Panel B: First-Stage Regressions for Expropriation Risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(
Log Settler Mortality	-0.417*	-0.376+	-0.336+	-0.302+	-0.356+	-0.291	-0.409*	-0.399*	-0.399*	
	(0.186)	(0.188)	(0.172)	(0.178)	(0.179)	(0.187)	(0.169)	(0.172)	(0.172)	
Yellow Fever Dummy										-1.(
										(0
Latitude (abs)							-0.811	-0.836	-0.836	
							(1.796)	(1.812)	(1.812)	
Coastal Dummy							0.574	0.546	0.546	
							(0.507)	(0.516)	(0.516)	
Mean Temperature							-0.116*	-0.118*	-0.118*	
							(0.052)	(0.053)	(0.053)	
Latitude		1.684		1.134		1.587				
		(1.398)		(1.397)		(1.379)				
Malaria Index	-0.790	-0.649								
	(0.538)	(0.548)								

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(
Life Expectancy			0.047*	0.044*						
			(0.019)	(0.019)						
Infant Mortality					-0.013*	-0.012*				
					(0.006)	(0.006)				
Africa Dummy										
Asia Dummy										
Constant	8.757***	8.209***	5.139**	4.987**	8.892***	8.270***	11.063***	11.054***	11.054***	7.3
	(0.751)	(0.876)	(1.808)	(1.823)	(0.688)	(0.873)	(1.421)	(1.433)	(1.433)	(0
Num. Obs.	62	62	60	60	60	60	60	59	59	
R-squared	0.295	0.312	0.341	0.349	0.323	0.339	0.372	0.360	0.360	0

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# Panel C: OLS regressions (Table 7 Panel C)

modelsummary(
ols_models_p7,
output = "gt",
title = "Table 7, Panel C: OLS Regressions with Geography and Health Variables",
coef_map = c("avexpr" = "Average Expropriation Risk", "lat_abst" = "Latitude",
"malfal94" = "Malaria Index", "leb95" = "Life Expectancy",
"imr95" = "Infant Mortality", "africa" = "Africa Dummy",
"asia" = "Asia Dummy", "other_cont" = "Other Continent Dummy"),
gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
stars = TRUE
)
```

Table 7, Panel C: OLS Regressions with Geography and Health Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Average Expropriation Risk	0.351***	0.346***	0.281***	0.279***	0.289***	0.277***	0.344***	0.286***	0.293***
	(0.055)	(0.057)	(0.052)	(0.053)	(0.050)	(0.051)	(0.053)	(0.051)	(0.049)

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Latitude		0.258		0.147		0.534			
		(0.629)		(0.556)		(0.527)			
Malaria Index	-1.135***	-1.108***					-1.127***		
	(0.190)	(0.203)					(0.181)		
Life Expectancy			0.052***	0.051***				0.050***	
			(0.007)	(0.007)				(0.007)	
Infant Mortality					-0.016***	-0.015***			-0.015***
					(0.002)	(0.002)			(0.002)
Africa Dummy									
Asia Dummy									
Num. Obs.	62	62	60	60	60	60	60	59	59
R-squared	0.713	0.714	0.770	0.770	0.778	0.782	0.732	0.765	0.774

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# --- Load and Prepare Data for Part 9 ---

ajr_dta8 <- read_dta("~/Desktop/maketable8.dta")
base_sample8 <- ajr_dta8 %>% filter(baseco == 1)
```

```
# Panel A & B: alternative instruments

panel_AB_models <- list(
"(1)" = feols(logpgp95 ~ 1 | avexpr ~ euro1900, data = base_sample8),
"(2)" = feols(logpgp95 ~ lat_abst | avexpr ~ euro1900, data = base_sample8),
"(3)" = feols(logpgp95 ~ 1 | avexpr ~ cons00a, data = base_sample8),
"(4)" = feols(logpgp95 ~ lat_abst | avexpr ~ cons00a, data = base_sample8),
"(5)" = feols(logpgp95 ~ 1 | avexpr ~ democ00a, data = base_sample8),
"(6)" = feols(logpgp95 ~ lat_abst | avexpr ~ democ00a, data = base_sample8),
"(7)" = feols(logpgp95 ~ indtime | avexpr ~ cons1, data = base_sample8),
"(8)" = feols(logpgp95 ~ lat_abst + indtime | avexpr ~ cons1, data = base_sample8),
"(9)" = feols(logpgp95 ~ indtime | avexpr ~ democ1, data = base_sample8),
"(10)" = feols(logpgp95 ~ lat_abst + indtime | avexpr ~ democ1, data = base_sample8)
)
```

```
## NOTE: 1 observation removed because of NA values (IV: 0/1).
## NOTE: 1 observation removed because of NA values (IV: 0/1).
```

```
## NOTE: 4 observations removed because of NA values (IV: 0/4).
## NOTE: 4 observations removed because of NA values (IV: 0/4).
```

```
## NOTE: 5 observations removed because of NA values (IV: 0/5).
## NOTE: 5 observations removed because of NA values (IV: 0/5).
```

```
## NOTE: 4 observations removed because of NA values (RHS: 4, IV: 0/4).
## NOTE: 4 observations removed because of NA values (RHS: 4, IV: 0/4).
```

```
## NOTE: 5 observations removed because of NA values (RHS: 4, IV: 0/5).
## NOTE: 5 observations removed because of NA values (RHS: 4, IV: 0/5).
```

```
# Panel C: overidentification tests (multiple instruments)

panel_C_models <- list(
"(1)" = feols(logpgp95 ~ 1 | avexpr ~ euro1900 + logem4, data = base_sample8),
"(2)" = feols(logpgp95 ~ lat_abst | avexpr ~ euro1900 + logem4, data = base_sample8),
"(3)" = feols(logpgp95 ~ 1 | avexpr ~ cons00a + logem4, data = base_sample8),
"(4)" = feols(logpgp95 ~ lat_abst | avexpr ~ cons00a + logem4, data = base_sample8),
"(5)" = feols(logpgp95 ~ 1 | avexpr ~ democ00a + logem4, data = base_sample8),
"(6)" = feols(logpgp95 ~ lat_abst | avexpr ~ democ00a + logem4, data = base_sample8),
"(7)" = feols(logpgp95 ~ indtime | avexpr ~ cons1 + logem4, data = base_sample8),
"(8)" = feols(logpgp95 ~ lat_abst + indtime | avexpr ~ cons1 + logem4, data = base_sample8),
"(9)" = feols(logpgp95 ~ indtime | avexpr ~ democ1 + logem4, data = base_sample8),
"(10)" = feols(logpgp95 ~ lat_abst + indtime | avexpr ~ democ1 + logem4, data = base_sample8)
)
```

```
## NOTE: 1 observation removed because of NA values (IV: 0/1).
```

```
## NOTE: 1 observation removed because of NA values (IV: 0/1).
```

```
## NOTE: 4 observations removed because of NA values (IV: 0/4).
## NOTE: 4 observations removed because of NA values (IV: 0/4).
```

```
## NOTE: 5 observations removed because of NA values (IV: 0/5).
## NOTE: 5 observations removed because of NA values (IV: 0/5).
```

```
## NOTE: 4 observations removed because of NA values (RHS: 4, IV: 0/4).
## NOTE: 4 observations removed because of NA values (RHS: 4, IV: 0/4).
```

```
## NOTE: 5 observations removed because of NA values (RHS: 4, IV: 0/5).
## NOTE: 5 observations removed because of NA values (RHS: 4, IV: 0/5).
```

```
# Panel D: IV with mortality as exogenous control

panel_D_models <- list(
"(1)" = feols(logpgp95 ~ logem4 | avexpr ~ euro1900, data = base_sample8),
"(2)" = feols(logpgp95 ~ logem4 + lat_abst | avexpr ~ euro1900, data = base_sample8),
"(3)" = feols(logpgp95 ~ logem4 | avexpr ~ cons00a, data = base_sample8),
"(4)" = feols(logpgp95 ~ logem4 + lat_abst | avexpr ~ cons00a, data = base_sample8),
"(5)" = feols(logpgp95 ~ logem4 | avexpr ~ democ00a, data = base_sample8),
"(6)" = feols(logpgp95 ~ logem4 + lat_abst | avexpr ~ democ00a, data = base_sample8),
"(7)" = feols(logpgp95 ~ logem4 + indtime | avexpr ~ cons1, data = base_sample8),
"(8)" = feols(logpgp95 ~ logem4 + lat_abst + indtime | avexpr ~ cons1, data = base_sample8),
"(9)" = feols(logpgp95 ~ logem4 + indtime | avexpr ~ democ1, data = base_sample8),
"(10)" = feols(logpgp95 ~ logem4 + lat_abst + indtime | avexpr ~ democ1, data = base_sample8)
)
```

```
## NOTE: 1 observation removed because of NA values (IV: 0/1).
```

```
## NOTE: 1 observation removed because of NA values (IV: 0/1).
```

```
## NOTE: 4 observations removed because of NA values (IV: 0/4).
## NOTE: 4 observations removed because of NA values (IV: 0/4).
```

```
## NOTE: 5 observations removed because of NA values (IV: 0/5).
## NOTE: 5 observations removed because of NA values (IV: 0/5).
```

```
## NOTE: 4 observations removed because of NA values (RHS: 4, IV: 0/4).
## NOTE: 4 observations removed because of NA values (RHS: 4, IV: 0/4).
```

```
## NOTE: 5 observations removed because of NA values (RHS: 4, IV: 0/5).
## NOTE: 5 observations removed because of NA values (RHS: 4, IV: 0/5).
```

```
# Panel A: 2SLS with alternative instruments (Table 8, Panel A)

modelsummary(
panel_AB_models,
output = "gt",
title = "Table 8, Panel A: 2SLS with Alternative Instruments",
coef_map = c("fit_avexpr" = "Average Expropriation Risk", "lat_abst" = "Latitude", "indtime" = "Da
te of Independence"),
gof_map = "nobs",
stars = TRUE
)
```

Table 8, Panel A: 2SLS with Alternative Instruments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(1
Average Expropriation Risk	0.870***	0.917***	0.706***	0.677**	0.719***	0.690***	0.595***	0.611***	0.549***	0.5

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(1
	(0.139)	(0.200)	(0.145)	(0.197)	(0.136)	(0.186)	(0.144)	(0.169)	(0.119)	(0.1
Latitude		-0.474		0.339		0.312		-0.405		-0.
		(1.241)		(1.082)		(1.046)		(0.922)		(0.8
Date of Independence							0.005**	0.005***	0.005***	0.00
							(0.001)	(0.001)	(0.001)	(0.(
Num.Obs.	63	63	60	60	59	59	60	60	59	5

```
# Panel B: first-stage for alternative instruments

first_stage_AB <- purrr::map(panel_AB_models, ~.$iv_first_stage$avexpr)

modelsummary(
first_stage_AB,
output = "gt",
title = "Table 8, Panel B: First Stage for Alternative Instruments",
coef_map = c("euro1900" = "European Settlements in 1900", "cons00a" = "Constraint on Executive in
1900",
"democ00a" = "Democracy in 1900", "cons1" = "Constraint on Executive at Independence",
"democ1" = "Democracy at Independence", "lat_abst" = "Latitude", "indtime" = "Date of Independenc
e",
"(Intercept)" = "Constant"),
gof_map = list(list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "r.squared", "clean" = "R-squared", "fmt" = 3)),
stars = TRUE
)
```

Table 8, Panel B: First Stage for Alternative Instruments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
European Settlements in 1900	0.032***	0.029***							
	(0.006)	(0.008)							
Constraint on Executive in 1900			0.318***	0.255**					
			(0.083)	(0.090)					
Democracy in 1900					0.242***	0.203**			
					(0.057)	(0.067)			

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
Constraint on Executive at Independence							0.249**	0.220**		
							(0.079)	(0.078)		
Democracy at Independence									0.191***	0
									(0.049)	(
Latitude		0.764		2.346		1.727		2.949*		2
		(1.540)		(1.435)		(1.492)		(1.446)		(
Date of Independence							0.008**	0.005+	0.008**	0
							(0.003)	(0.003)	(0.003)	(
Constant	5.993***	5.896***	5.756***	5.484***	6.099***	5.858***	4.890***	4.722***	5.241***	5.
	(0.187)	(0.271)	(0.254)	(0.301)	(0.195)	(0.285)	(0.469)	(0.464)	(0.352)	(
Num. Obs.	63	63	60	60	59	59	60	60	59	
R-squared	0.299	0.301	0.203	0.239	0.238	0.255	0.190	0.246	0.257	

```
# Panel C: Overidentification tests — include Sargan p-value if modelsummary supports it for your
feols objects

gof_map_sargan <- list(
list("raw" = "nobs", "clean" = "Num. Obs.", "fmt" = 0),
list("raw" = "iv_sargan_p", "clean" = "Sargan Test p-value", "fmt" = 3)
)
modelsummary(
panel_C_models,
output = "gt",
title = "Table 8, Panel C: Overidentification Tests",
coef_map = c("fit_avexpr" = "Average Expropriation Risk", "lat_abst" = "Latitude", "indtime" = "Da
te of Independence"),
gof_map = gof_map_sargan,
stars = TRUE
)
```

Table 8, Panel C: Overidentification Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
Average Expropriation Risk	0.893***	0.946***	0.808***	0.833***	0.799***	0.820***	0.670***	0.705***	0.633***	0.

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
	(0.128)	(0.173)	(0.132)	(0.177)	(0.127)	(0.167)	(0.117)	(0.142)	(0.108)	(
Latitude		-0.597		-0.299		-0.214		-0.752		-
		(1.186)		(1.109)		(1.065)		(0.914)		(
Date of Independence							0.005**	0.005**	0.005**	0
							(0.001)	(0.002)	(0.001)	(
Num. Obs.	63	63	60	60	59	59	60	60	59	

```
# Panel D: 2SLS with log mortality as exogenous control

modelsummary(
panel_D_models,
output = "gt",
title = "Table 8, Panel D: 2SLS with Log Mortality as Exogenous Variable",
coef_map = c("fit_avexpr" = "Average Expropriation Risk", "logem4" = "Log Settler Mortality",
"lat_abst" = "Latitude", "indtime" = "Date of Independence"),
gof_map = "nobs",
stars = TRUE
)
```

Table 8, Panel D: 2SLS with Log Mortality as Exogenous Variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average Expropriation Risk	0.814***	0.879**	0.454+	0.416	0.515*	0.481+	0.485*	0.494+	0.402*	0.407*
	(0.229)	(0.292)	(0.249)	(0.299)	(0.227)	(0.275)	(0.232)	(0.248)	(0.180)	(0.191)
Log Settler Mortality	-0.067	-0.050	-0.253	-0.259	-0.214	-0.222	-0.136	-0.143	-0.187	-0.190
	(0.165)	(0.183)	(0.164)	(0.169)	(0.154)	(0.161)	(0.155)	(0.149)	(0.125)	(0.122)
Latitude		-0.522		0.383		0.278		-0.375		-0.167
		(1.150)		(0.895)		(0.863)		(0.842)		(0.733)
Date of Independence							0.004***	0.005**	0.004***	0.004**
							(0.001)	(0.001)	(0.001)	(0.001)
Num.Obs.	63	63	60	60	59	59	60	60	59	59

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001