Replication of Comin, Easterly, and Gong (2010)

```
packages = c("dplyr", "readr", "haven", "estimatr",
             "conleyreg", "lmtest", "sandwich", "tibble",
             "quantreg", "stargazer")
sapply(packages, library, character.only = TRUE)
```

Load in the data:

```
wb_data = read_csv("../datasets/world-regions-according-to-the-world-bank.csv")
macro_data = read_dta("../datasets/2-comin-easterly-gong/primitive_aejmacro.dta")
```

Merge with World Bank regions:

```
macro_strip = lapply(macro_data, function(col) {
  attr(col, "format.stata") <- NULL
  return(col)
})

macro_strip = as.data.frame(macro_strip)

wb_regions = wb_data |>
  janitor::clean_names() |>
  select(-year, -code) |>
  rename(wb_region = world_region_according_to_the_world_bank,
         country = entity)

macro_final = left_join(macro_strip, wb_regions,
                        by = "country") |>
  mutate(wb_region = case_when(
    country %in% c("French Guiana", "Lesser Antilles") ~ "Latin American and Caribbean",
```

```
    country %in% c("Democratic Republic of the Congo", "Republic of the Congo",
                   "The Gambia", "Swaziland", "Cote D'Ivoire") ~ "Sub-Saharan Africa",
    country %in% c("Czechoslovakia", "Macedonia", "Serbia & Montenegro",
                   "England") ~ "Europe and Central Asia",
    country %in% c("Cook Islands", "Korea") ~ "East Asia and Pacific",
    TRUE ~ wb_region
  )) |>
  mutate(mena = as.integer(wb_region == "Middle East and North Africa"),
         eur = as.integer(wb_region == "Europe and Central Asia"),
         sa = as.integer(wb_region == "South Asia"),
         eap = as.integer(wb_region == "East Asia and Pacific"),
         na = as.integer(wb_region == "North America"),
         latam = as.integer(wb_region == "Latin America and Caribbean"),
         afr = as.integer(wb_region == "Sub-Saharan Africa"))
```

The following code replicates Column 1 of Table 8A (impact of technology adoption in 1000 BC on log per capita income in 2002). The first equation follows Comin et al's regression, while the second follows Kelly and introduces controls for World Bank regions. In both cases, standard errors are clustered.

```
clus1000 = macro_final$clus1000

mod1 = lm(ly2002 ~ tr1,
                 data = macro_final)
mod1_se = vcovCL(mod1, cluster = clus1000)
rob1 = coeftest(mod1, vcov = mod1_se)[, "Std. Error"]

mod2 = lm(ly2002 ~ tr1 + mena + eur + sa + eap + na + latam,
                 data = macro_final)
mod2_se = vcovCL(mod2, cluster = clus1000)
rob2 = coeftest(mod2, vcov = mod2_se)[, "Std. Error"]
```

This chunk replicates Column 3 of Table 8A (impact of technology adoption in 1500 AD on log per capita income in 2002).

```
clusc = na.omit(macro_final$clusc)

mod3 = lm(ly2002 ~ tr1500cc,
          data = macro_final |> filter(!is.na(clusc)))
mod3_se = vcovCL(mod3, cluster = clusc)
rob3 = coeftest(mod3, vcov = mod3_se)[, "Std. Error"]
```

```
mod4 = lm(ly2002 ~ tr1500cc + mena + eur + sa + eap + na + latam,
                  data = macro_final |> filter(!is.na(clusc)))
mod4_se = vcovCL(mod4, cluster = clusc)
rob4 = coeftest(mod4, vcov = mod4_se)[, "Std. Error"]

#output = stargazer(mod1, mod2, mod3, mod4,
#          se = list(rob1, rob2, rob3, rob4),
#          type = "latex",
#          dep.var.labels = "Log income per capita in 2002",
#          covariate.labels = c("Overall technology adoption level in 1000 BC",
#                               "MENA dummy", "Europe dummy", "South Asia dummy",
#                               "East Asia dummy", "North American dummy",
#                               "Latin America dummy",
#                               "Overall technology adoption level in 1500 AD"),
#          star.cutoffs = c(0.05, 0.01, 0.001),
#          star.char = c("*", "**", "***"),
#          omit.stat = c("f", "ser"))
```

In both cases, adding the World Bank region dummies reduces the effect size significantly. In the case of technology adoption in 1000 BC, the effect reverses direction. In the original paper, the authors control for continental variation in Table 10. I tried to replicate these equations, but got different effects. I'm not sure how to account for these differences.

```
mod5 = lm(ly2002 ~ tr1 + eu + af + as + am,
          data = macro_final)
mod5_se = vcovCL(mod5, cluster = clus1000)
coeftest(mod5, vcov = mod5_se)
```

```
t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.024724   0.919414  9.8157 3.04e-16 ***
tr1         -0.061553   0.696192 -0.0884   0.9297
eu           0.602219   0.969788  0.6210   0.5361
af          -1.461469   0.948083 -1.5415   0.1264
as          -0.702623   0.944212 -0.7441   0.4586
am          -0.232663   0.922948 -0.2521   0.8015
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod6 = lm(ly2002 ~ tr2 + eu + af + as + am,
          data = macro_final)
mod6_se = vcovCL(mod6, cluster = clus1000)
coeftest(mod6, vcov = mod6_se)
```

```
t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.29966    0.78116 11.9050   <2e-16 ***
tr2          0.22992    0.47147  0.4877   0.6267
eu           0.11672    0.92867  0.1257   0.9002
af          -2.09652    0.88652 -2.3649   0.0197 *
as          -1.06788    0.88132 -1.2117   0.2281
am          -0.59970    0.80785 -0.7423   0.4594
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod7 = lm(ly2002 ~ tr1500cc + eu + af + as + am,
          data = macro_final |> filter(!is.na(clusc)))
mod7_se = vcovCL(mod7, cluster = clusc)
coeftest(mod7, vcov = mod7_se)
```

```
t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  8.95986    0.56240 15.9314 < 2.2e-16 ***
tr1500cc     1.26590    0.87714  1.4432  0.151936
eu          -0.41371    0.88447 -0.4677  0.640934
af          -1.94682    0.66209 -2.9404  0.004032 **
as          -1.38743    0.75737 -1.8319  0.069802 .
am          -0.36602    0.58430 -0.6264  0.532391
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The following code adds the World Bank region dummies to the migration-adjusted technology measure regressions.

```r
mod8 = lm(ly2002 ~ tr1mig + mena + eur + sa + eap + na + latam,
          data = macro_final)
mod8_se = vcovCL(mod8, cluster = clus1000)
rob8 = coeftest(mod8, vcov = mod8_se)[, "Std. Error"]

mod9 = lm(ly2002 ~ tr2mig + mena + eur + sa + eap + na + latam,
          data = macro_final)
mod9_se = vcovCL(mod9, cluster = clus1000)
rob9 = coeftest(mod9, vcov = mod9_se)[, "Std. Error"]

mod10 = lm(ly2002 ~ t3mig + mena + eur + sa + eap + na + latam,
          data = macro_final |> filter(!is.na(clusc)))
mod10_se = vcovCL(mod10, cluster = clusc)
rob10 = coeftest(mod10, vcov = mod10_se)[, "Std. Error"]
```