Data Management & Analysis Final Project

Replication and Extention for Acemoglu, Naidu, Restrepo and Robinson (2019)

(Name:) Shoya Abe (University ID:) 31B24001
(Name:) Honoka Ohtani (University ID:) 31B24002

(Submission Due:) 2025/02/06

Contents

0.1 Setup

```r
pacman::p_load(
  rmdformats,
  knitr,
  tinytex,
  haven,
  tidyverse,
  kableExtra,
  plm,
  texreg
)

## Global options
options(max.print="75")
opts_chunk$set(fig.align="center",
               echo=TRUE,
               cache=TRUE,
               prompt=FALSE,
               tidy=FALSE,
               comment=NA,
               message=FALSE,
               warning=FALSE)
opts_knit$set(width=75)
```

1 About this Report

1.1 Project Type

1.2 Summary of the Paper

1.2.1 What the problem is

1.2.2 Why it is important

1.2.3 How you solve the problem

1.2.4 What we find

1.3 Data

```r
data <- read_dta("data/raw/DDCGdata_final.dta")

# Define the function to summarize the data
summarize_data <- function(data, n = 10) {
  cat("Sample size (number of rows):",
      nrow(data), "\n")
  cat("Number of variables (columns):",
      ncol(data), "\n")
```

```r
  cat("Variable names (first",
      n, "names):\n")
  print(head(colnames(data), n))
}

summarize_data(data)
```

```
Sample size (number of rows): 9384
Number of variables (columns): 1177
Variable names (first 10 names):
 [1] "country_name"                "wbcode"
 [3] "year"                        "gdppercapitaconstant2000us"
 [5] "lp_bl"                       "ls_bl"
 [7] "lh_bl"                       "taxratio"
 [9] "region"                      "wbcode2"
```

1.4 Empirical Methods

1.4.1 Event Study (Figure.1)

1.4.2 Dynamic Liner Panel Model (Table.2)

2 Replication

2.1 Figure.1

2.1.1 Preprocessing

```r
data_f1 <- data %>%
  rename(id = "_ID") %>%
  group_by(id) %>%
  arrange(year) %>%
  mutate(
    prev_dem = dplyr::lag(dem, 1),
    transition = case_when(
      dem == 1 & prev_dem == 0 ~ 1,
      dem == 0 & prev_dem == 0 ~ 0,
      TRUE ~ NA_real_
    ),
    lag1 = dplyr::lag(y, 1),
    lag2 = dplyr::lag(y, 2),
    lag3 = dplyr::lag(y, 3),
    lag4 = dplyr::lag(y, 4)
  ) %>%
  filter(
    !is.na(lag1) &
      !is.na(lag2) &
      !is.na(lag3) &
      !is.na(lag4)
  ) %>%
```

```
  ungroup()

for (t in -15:-2) {
  col_name <- paste0("gdpDiff_m", abs(t))
  data_f1 <- data_f1 %>%
    group_by(id) %>%
    arrange(year) %>%
    mutate(!!col_name := dplyr::lag(y, abs(t)) - lag1) %>%
    ungroup()
}

data_f1 <- data_f1 %>%
  mutate(
    gdpDiff_m1 = 0,
    gdpDiff_0  = y - lag1
  )

for (t in 1:30) {
  col_name <- paste0("gdpDiff_p", t)
  data_f1 <- data_f1 %>%
    group_by(id) %>%
    arrange(year) %>%
    mutate(!!col_name := dplyr::lead(y, t) - lag1) %>%
    ungroup()
}

data_f1 <- data_f1 %>%
  filter(!is.na(transition))
```

2.1.2 Estimation

```
estimateATET <- function(outcome_col) {
  sub_data <- data_f1 %>%
    filter(
      !is.na(.data[[outcome_col]]),
      !is.na(transition)
    )
  if (nrow(sub_data) == 0)
    return(NA)

  year_levels <- sort(unique(sub_data$year))
  sub_data <- sub_data %>%
    mutate(year_factor = factor(year, levels = year_levels))

  control_data <- sub_data %>%
    filter(transition == 0)
  treated_data <- sub_data %>%
    filter(transition == 1)

  if (nrow(control_data) < 2 ||
      length(unique(control_data$year)) < 2)
    return(NA)
```

```r
  model_formula <- as.formula(
    paste(outcome_col, "~ year_factor - 1")
  )

  control_model <- tryCatch(
    lm(model_formula, data = control_data),
    error = function(e) NULL
  )

  if (is.null(control_model))
    return(NA)

  predicted_outcomes <- tryCatch(
    predict(control_model, newdata = treated_data),
    error = function(e) rep(NA, nrow(treated_data))
  )

  treatment_effects <- treated_data[[outcome_col]] - predicted_outcomes
  mean(treatment_effects, na.rm = TRUE)
}

relative_times <- c(seq(-15, -1), seq(0, 30))
atets <- numeric(length(relative_times))

for (i in seq_along(relative_times)) {
  t_val <- relative_times[i]
  if (t_val < 0) {
    col_name <- paste0("gdpDiff_m", abs(t_val))
  } else {
    col_name <- if (t_val == 0)
      "gdpDiff_0" else paste0("gdpDiff_p", t_val)
  }
  atets[i] <- estimateATET(col_name)
}

results_df <- data.frame(
  RelativeTime = relative_times,
  ATET         = atets
)
```

2.1.3 Plot

```r
figure_1 <- ggplot(results_df, aes(x = RelativeTime, y = ATET)) +
  geom_line(color = "black") +
  scale_x_continuous(breaks = seq(-15, 30, 5)) +
  labs(
    x = "Years around Democratization",
    y = "Change in GDP per capita (log points)"
  ) +
  theme_bw()
```

```r
ggsave("output/figure_1.pdf",
       width = 14,
       height = 8,
       units = "cm")
```

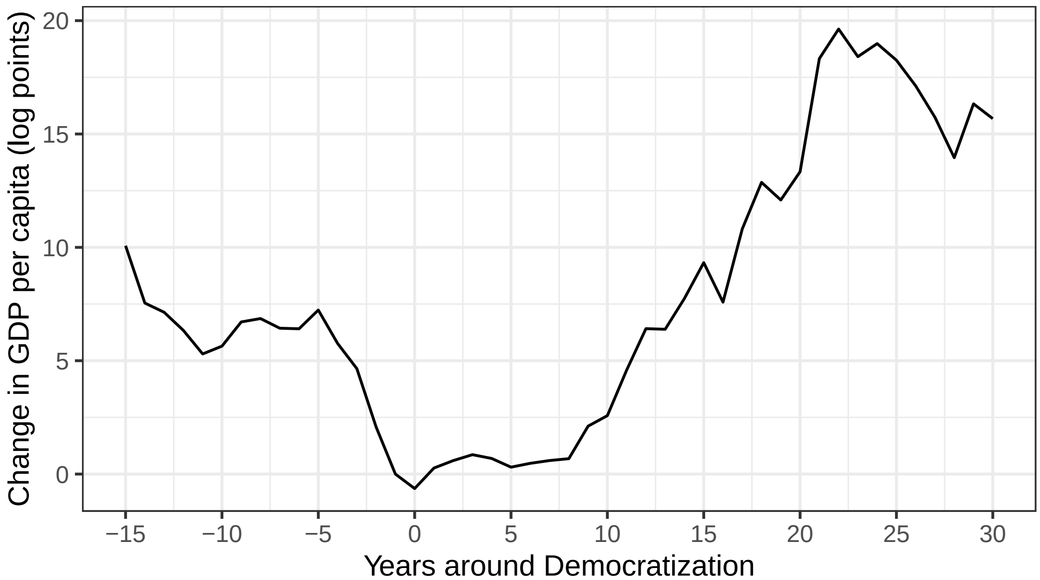

Figure 1: Event Study

2.2 Table.1

2.2.1 Preprocessing

```r
var_info <- tibble(
  var = c(
    "gdppercapitaconstant2000us",
    "ginv",
    "tradewb",
    "prienr",
    "secenr",
    "taxratio",
    "mortnew",
    "unrestn",
    "marketref"
  ),
  label = c(
    "GDP per capita",
    "Investment share of GDP",
    "Trade share of GDP",
    "Primary-school enrollment rate",
    "Secondary-school enrollment rate",
    "Tax revenue share of GDP",
```

```
    "Child mortality per 1,000 births",
    "Unrest rate",
    "Market reforms index (0-100)"
  )
)

calc_summary <- function(df, var, group_var) {
  df %>%
    filter(!is.na(.data[[var]])) %>%
    group_by({{ group_var }}) %>%
    summarise(
      Observations = n(),
      Mean         = mean(.data[[var]], na.rm = TRUE),
      SD           = sd(.data[[var]], na.rm = TRUE),
      .groups      = "drop"
    ) %>%
    mutate(Variable = var)
}

var_list <- var_info$var
```

2.2.2 Caliculation

```
summary_table <- lapply(
  var_list,
  function(x) calc_summary(data, x, dem)
) %>%
  bind_rows() %>%
  pivot_wider(
    names_from  = dem,
    values_from = c(Observations, Mean, SD),
    names_glue  = "{.value}_dem{dem}"
  ) %>%
  rename(
    Observations_Nondem = Observations_dem0,
    Mean_Nondem         = Mean_dem0,
    SD_Nondem           = SD_dem0,
    Observations_Dem    = Observations_dem1,
    Mean_Dem            = Mean_dem1,
    SD_Dem              = SD_dem1
  ) %>%
  left_join(
    var_info,
    by = c("Variable" = "var")
  ) %>%
  select(
    label,
    Observations_Nondem,
    Mean_Nondem,
    SD_Nondem,
    Observations_Dem,
    Mean_Dem,
```

```
    SD_Dem
  ) %>%
  rename(Variable = label)

colnames(summary_table) <- c(
  "Variable",
  "Observations",
  "Mean",
  "SD",
  "Observations",
  "Mean",
  "SD"
)
```

2.2.3 Tabulation

```
latex_table <- summary_table %>%
  kbl(
    caption   = "Summary Statistics by Democracy Status",
    format    = "latex",
    booktabs  = TRUE,
    digits    = 2
  ) %>%
  add_header_above(
    c(" " = 1, "Nondemocracies" = 3, "Democracies" = 3)
  ) %>%
  kable_styling(
    latex_options = c("HOLD_position", "striped")
  )

save_kable(latex_table, file = "output/table_1.tex")
```

Table 1: Summary Statistics by Democracy Status

	Nondemocracies			Democracies		
Variable	Observations	Mean	SD	Observations	Mean	SD
GDP per capita	3376	2074.46	3838.65	3558	8149.97	9334.83
Investment share of GDP	3225	21.82	10.23	3340	23.28	7.41
Trade share of GDP	3175	71.63	51.06	3485	77.15	41.04
Primary-school enrollment rate	2861	90.29	29.51	2823	101.60	15.86
Secondary-school enrollment rate	2424	45.76	31.77	2538	75.40	29.78
Tax revenue share of GDP	3122	0.16	0.09	2564	0.21	0.10
Child mortality per 1,000 births	4142	77.29	49.64	3615	33.26	32.65
Unrest rate	3739	28.70	45.24	3610	21.91	41.37
Market reforms index (0–100)	3476	21.89	23.26	2829	52.11	24.75

2.3 Table.2

2.3.1 Preprocessing

```r
data_t2 <- data %>%
  select(1:30) %>%
  group_by(country_name) %>%
  arrange(year) %>%
  mutate(
    lag1 = dplyr::lag(y, 1),
    lag2 = dplyr::lag(y, 2),
    lag3 = dplyr::lag(y, 3),
    lag4 = dplyr::lag(y, 4),
    lag5 = dplyr::lag(y, 5),
    lag6 = dplyr::lag(y, 6),
    lag7 = dplyr::lag(y, 7),
    lag8 = dplyr::lag(y, 8)
  ) %>%
  ungroup()
```

2.3.2 Estimation

```r
data_m1 <- data_t2 %>%
  drop_na(y, dem, lag1)
data_m1 <- pdata.frame(data_m1,
                       index = c("country_name", "year"))
model_1 <- plm(
  y ~ dem + lag1,
  data   = data_m1,
  model  = "within",
  effect = "twoways"
)

data_m2 <- data_t2 %>%
  drop_na(y, dem, lag1, lag2)
data_m2 <- pdata.frame(data_m2,
                       index = c("country_name", "year"))
model_2 <- plm(
  y ~ dem + lag1 + lag2,
  data   = data_m2,
  model  = "within",
  effect = "twoways"
)

data_m3 <- data_t2 %>%
  drop_na(y, dem, lag1, lag2, lag3, lag4)
data_m3 <- pdata.frame(data_m3,
                       index = c("country_name", "year"))
model_3 <- plm(
  y ~ dem + lag1 + lag2 + lag3 + lag4,
  data   = data_m3,
```

```r
  model  = "within",
  effect = "twoways"
)

data_m4 <- data_t2 %>%
  drop_na(y, dem, lag1, lag2, lag3, lag4,
          lag5, lag6, lag7, lag8)
data_m4 <- pdata.frame(data_m4,
                       index = c("country_name", "year"))
model_4 <- plm(
  y ~ dem + lag1 + lag2 + lag3 + lag4 +
    lag5 + lag6 + lag7 + lag8,
  data   = data_m4,
  model  = "within",
  effect = "twoways"
)

beta_hat_1  <- coef(model_1)["dem"]
gamma_hat_1 <- coef(model_1)[c("lag1")]
long_run_effect_1 <- beta_hat_1 / (1 - sum(gamma_hat_1))

beta_hat_2  <- coef(model_2)["dem"]
gamma_hat_2 <- coef(model_2)[c("lag1", "lag2")]
long_run_effect_2 <- beta_hat_2 / (1 - sum(gamma_hat_2))

beta_hat_3  <- coef(model_3)["dem"]
gamma_hat_3 <- coef(model_3)[c("lag1", "lag2", "lag3", "lag4")]
long_run_effect_3 <- beta_hat_3 / (1 - sum(gamma_hat_3))

beta_hat_4  <- coef(model_4)["dem"]
gamma_hat_4 <- coef(model_4)[c("lag1", "lag2", "lag3",
                               "lag4", "lag5", "lag6",
                               "lag7", "lag8")]
long_run_effect_4 <- beta_hat_4 / (1 - sum(gamma_hat_4))

lre <- round(
  c(
    long_run_effect_1,
    long_run_effect_2,
    long_run_effect_3,
    long_run_effect_4
  ),
  3
)
print(lre)
```

```
   dem    dem    dem    dem
35.587 19.599 21.240 22.008
```

```r
pers1 <- sum(coef(model_1)[2])
pers2 <- sum(coef(model_2)[2:3])
pers3 <- sum(coef(model_3)[2:5])
pers4 <- sum(coef(model_4)[2:9])
```

```r
pers <- round(
  c(pers1, pers2, pers3, pers4),
  3
)
print(pers)
```

```
[1] 0.973 0.967 0.963 0.960
```

```r
dem_shortrun <- coef(model_1)["dem"]
lag1_mod1 <- coef(model_1)[2]
effect1 <- dem_shortrun
effect2 <- (effect1 * lag1_mod1) + dem_shortrun
effects_mod1 <- c(effect1, effect2)

for (i in 3:30) {
  eff <- (effects_mod1[i-1] * lag1_mod1) + dem_shortrun
  effects_mod1 <- c(effects_mod1, eff)
}
eff_25_1 <- effects_mod1[25]

dem_shortrun <- coef(model_2)["dem"]
lag1_mod2 <- coef(model_2)[2]
lag2_mod2 <- coef(model_2)[3]
effect1 <- dem_shortrun
effect2 <- (effect1 * lag1_mod2) + dem_shortrun
effect3 <- (effect2 * lag1_mod2) +
  (effect1 * lag2_mod2) +
  dem_shortrun
effects_mod2 <- c(effect1, effect2, effect3)

for (i in 4:30) {
  eff <- (effects_mod2[i-1] * lag1_mod2) +
    (effects_mod2[i-2] * lag2_mod2) +
    dem_shortrun
  effects_mod2 <- c(effects_mod2, eff)
}
eff_25_2 <- effects_mod2[25]

dem_shortrun <- coef(model_3)["dem"]
lag1_mod3 <- coef(model_3)[2]
lag2_mod3 <- coef(model_3)[3]
lag3_mod3 <- coef(model_3)[4]
lag4_mod3 <- coef(model_3)[5]
effect1 <- dem_shortrun
effect2 <- (effect1 * lag1_mod3) + dem_shortrun
effect3 <- (effect2 * lag1_mod3) +
  (effect1 * lag2_mod3) +
  dem_shortrun
effect4 <- (effect3 * lag1_mod3) +
  (effect2 * lag2_mod3) +
  (effect1 * lag3_mod3) +
  dem_shortrun
effects_mod3 <- c(effect1, effect2, effect3, effect4)
```

```
for (i in 5:30) {
  eff <- (effects_mod3[i-1] * lag1_mod3) +
    (effects_mod3[i-2] * lag2_mod3) +
    (effects_mod3[i-3] * lag3_mod3) +
    (effects_mod3[i-4] * lag4_mod3) +
    dem_shortrun
  effects_mod3 <- c(effects_mod3, eff)
}
eff_25_3 <- effects_mod3[25]

dem_shortrun <- coef(model_4)["dem"]
lag1_mod4 <- coef(model_4)[2]
lag2_mod4 <- coef(model_4)[3]
lag3_mod4 <- coef(model_4)[4]
lag4_mod4 <- coef(model_4)[5]
lag5_mod4 <- coef(model_4)[6]
lag6_mod4 <- coef(model_4)[7]
lag7_mod4 <- coef(model_4)[8]
lag8_mod4 <- coef(model_4)[9]
effect1 <- dem_shortrun
effect2 <- (effect1 * lag1_mod4) + dem_shortrun
effect3 <- (effect2 * lag1_mod4) +
  (effect1 * lag2_mod4) +
  dem_shortrun
effect4 <- (effect3 * lag1_mod4) +
  (effect2 * lag2_mod4) +
  (effect1 * lag3_mod4) +
  dem_shortrun
effect5 <- (effect4 * lag1_mod4) +
  (effect3 * lag2_mod4) +
  (effect2 * lag3_mod4) +
  (effect1 * lag4_mod4) +
  dem_shortrun
effect6 <- (effect5 * lag1_mod4) +
  (effect4 * lag2_mod4) +
  (effect3 * lag3_mod4) +
  (effect2 * lag4_mod4) +
  (effect1 * lag5_mod4) +
  dem_shortrun
effect7 <- (effect6 * lag1_mod4) +
  (effect5 * lag2_mod4) +
  (effect4 * lag3_mod4) +
  (effect3 * lag4_mod4) +
  (effect2 * lag5_mod4) +
  (effect1 * lag6_mod4) +
  dem_shortrun
effect8 <- (effect7 * lag1_mod4) +
  (effect6 * lag2_mod4) +
  (effect5 * lag3_mod4) +
  (effect4 * lag4_mod4) +
  (effect3 * lag5_mod4) +
  (effect2 * lag6_mod4) +
  (effect1 * lag7_mod4) +
```

```
  dem_shortrun
effects_mod4 <- c(effect1, effect2, effect3, effect4,
                  effect5, effect6, effect7, effect8)

for (i in 9:30) {
  eff <- (effects_mod4[i-1] * lag1_mod4) +
    (effects_mod4[i-2] * lag2_mod4) +
    (effects_mod4[i-3] * lag3_mod4) +
    (effects_mod4[i-4] * lag4_mod4) +
    (effects_mod4[i-5] * lag5_mod4) +
    (effects_mod4[i-6] * lag6_mod4) +
    (effects_mod4[i-7] * lag7_mod4) +
    (effects_mod4[i-8] * lag8_mod4) +
    dem_shortrun
  effects_mod4 <- c(effects_mod4, eff)
}
eff_25_4 <- effects_mod4[25]

eff_25 <- round(
  c(eff_25_1, eff_25_2, eff_25_3, eff_25_4),
  3
)
print(eff_25)
```

```
   dem    dem    dem    dem
17.791 13.800 16.895 17.715
```

```
se1 <- sqrt(diag(vcov(model_1)))
se2 <- sqrt(diag(vcov(model_2)))
se3 <- sqrt(diag(vcov(model_3)))
se4 <- sqrt(diag(vcov(model_4)))

override.coef.1 <- c(
  coef(model_1)["dem"],
  coef(model_1)["lag1"],
  NA, NA, NA, NA, NA, NA, NA
)
override.se.1 <- c(
  se1["dem"],
  se1["lag1"],
  NA, NA, NA, NA, NA, NA, NA
)

override.coef.2 <- c(
  coef(model_2)["dem"],
  coef(model_2)["lag1"],
  coef(model_2)["lag2"],
  NA, NA, NA, NA, NA, NA
)
override.se.2 <- c(
  se2["dem"],
  se2["lag1"],
  se2["lag2"],
```

```
  NA, NA, NA, NA, NA, NA
)

override.coef.3 <- c(
  coef(model_3)["dem"],
  coef(model_3)["lag1"],
  coef(model_3)["lag2"],
  coef(model_3)["lag3"],
  coef(model_3)["lag4"],
  NA, NA, NA, NA
)
override.se.3 <- c(
  se3["dem"],
  se3["lag1"],
  se3["lag2"],
  se3["lag3"],
  se3["lag4"],
  NA, NA, NA, NA
)

override.coef.4 <- c(
  coef(model_4)["dem"],
  coef(model_4)["lag1"],
  coef(model_4)["lag2"],
  coef(model_4)["lag3"],
  coef(model_4)["lag4"],
  coef(model_4)["lag5"],
  coef(model_4)["lag6"],
  coef(model_4)["lag7"],
  coef(model_4)["lag8"]
)
override.se.4 <- c(
  se4["dem"],
  se4["lag1"],
  se4["lag2"],
  se4["lag3"],
  se4["lag4"],
  se4["lag5"],
  se4["lag6"],
  se4["lag7"],
  se4["lag8"]
)
```

2.3.3 Tabulation

```
models <- list(model_1, model_2, model_3, model_4)

texreg(
  models,
  override.coef = list(
    override.coef.1,
    override.coef.2,
```

```
    override.coef.3,
    override.coef.4
  ),
  override.se = list(
    override.se.1,
    override.se.2,
    override.se.3,
    override.se.4
  ),
  custom.model.names = c(
    "(1)", "(2)", "(3)", "(4)"
  ),
  custom.coef.names = c(
    "Democracy", "Lag 1", "Lag 2",
    "Lag 3", "Lag 4", "Lag 5",
    "Lag 6", "Lag 7", "Lag 8"
  ),
  custom.gof.rows = list(
    "Persistence"          = pers,
    "Long run effect"      = lre,
    "Effect after 25 years" = eff_25
  ),
  file = "output/table_2_FE.tex",
  caption = "Effect of Democracy on (Log) GDP per Capita"
)
```

2.4 Figure.2

2.5 Table.3

2.6 Table.4

2.7 Table.5

3 Extention

	(1)	(2)	(3)	(4)
Democracy	0.97^{***}	0.65^{**}	0.79^{***}	0.89^{***}
	(0.24)	(0.23)	(0.23)	(0.24)
Lag 1	0.97^{***}	1.27^{***}	1.24^{***}	1.23^{***}
	(0.00)	(0.01)	(0.01)	(0.01)
Lag 2		-0.30^{***}	-0.21^{***}	-0.21^{***}
		(0.01)	(0.02)	(0.02)
Lag 3			-0.03	-0.02
			(0.02)	(0.02)
Lag 4			-0.04^{***}	-0.04
			(0.01)	(0.02)
Lag 5				-0.02
				(0.02)
Lag 6				0.01
				(0.02)
Lag 7				0.02
				(0.02)
Lag 8				-0.01
				(0.01)
Persistence	0.97	0.97	0.96	0.96
Long run effect	35.59	19.60	21.24	22.01
Effect after 25 years	17.79	13.80	16.90	17.72
R^2	0.96	0.96	0.96	0.96
Adj. R^2	0.96	0.96	0.96	0.96
Num. obs.	6790	6642	6336	5688

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Effect of Democracy on (Log) GDP per Capita

	(1)	(2)	(3)	(4)
Democracy	2.79	2.29	0.05	1.51
	(2.12)	(1.63)	(1.42)	(0.51)
Lag 1	0.96^{***}	0.99^{***}	0.94^{***}	0.93^{***}
	(0.03)	(0.03)	(0.03)	(0.01)
Lag 2		-0.02	-0.00	-0.01
		(0.01)	(0.01)	(0.00)
Lag 3			0.00	0.00
			(0.01)	(0.00)
Lag 4			-0.02^{*}	-0.01
			(0.01)	(0.00)
Lag 5				-0.00
				(0.00)
Lag 6				0.00
				(0.00)
Lag 7				-0.00
				(0.00)
Lag 8				-0.00
				(0.00)
Persistence	0.96	0.97	0.92	0.91
Long run effect	63.18	74.26	0.65	16.40
Effect after 25 years	42.76	40.77	0.59	15.27
n	175	175	175	175
T	50	49	47	43
Num. obs.	6790	6642	6336	5688
Num. obs. used	6542	6311	5824	4779
Sargan Test: chisq	145.66	147.27	140.10	146.09
Sargan Test: df	2398.00	2297.00	2095.00	1691.00
Sargan Test: p-value	1.00	1.00	1.00	1.00
Wald Test Coefficients: chisq	808.19	984.51	1143.95	2227.71
Wald Test Coefficients: df	2	3	5	9
Wald Test Coefficients: p-value	0.00	0.00	0.00	0.00
Wald Test Time Dummies: chisq	533.24	491.67	497.42	453.37
Wald Test Time Dummies: df	48	46	42	34
Wald Test Time Dummies: p-value	0.00	0.00	0.00	0.00

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Effect of Democracy on (Log) GDP per Capita: Arellano–Bond GMM Estimation