Included Variable Bias in Rogowski et al. (2022): An Empirical Application

2026-02-27

1 Introduction

Rogowski, Gerring, Maguire, and Cojocaru (2022, AJPS) estimate the effect of public postal infrastructure on economic development using a two-way fixed effects (TWFE) country panel model. The main specification (Table 1, Model 2) is:

$$\text{GDP growth}_{i,t+1} = \beta \, \text{Post Offices}_{i,t-1} + \mathbf{Z}'_{it}\gamma + \delta_i + \gamma_t + \varepsilon_{it}$$

where GDP growth$_{i,t+1}$ is the 5-year GDP growth rate (one period forward/lead), Post Offices$_{i,t-1}$ is the log cumulative stock of post offices (5-year lag), $\mathbf{Z}_{it}$ is a vector of four controls (log GDP per capita, log population, urbanization rate, and Polity2 democracy score), δ_i are country fixed effects, and γ_t are year fixed effects. Standard errors are clustered by country.

A key question is whether any of these controls are **colliders** — variables caused by both the treatment (postal infrastructure) and the outcome (GDP growth). If so, conditioning on them introduces **Included Variable Bias** (IVB). The IVB for each candidate collider z_j is given by:

$$\text{IVB}_j(\hat{\beta}) = \hat{\beta}^{\text{long}} - \hat{\beta}^{\text{short}_j} = -\hat{\theta}_j \cdot \hat{\pi}_j$$

where $\hat{\theta}_j$ is the coefficient of z_j in the "long" model (with all controls) and $\hat{\pi}_j$ is the coefficient of the treatment in the auxiliary regression of z_j on the treatment and remaining controls.

We apply this decomposition to the Rogowski et al. data, treating each control as a candidate collider.

2 Data

```
dat <- read.delim(
  "country_panel.tab",
  stringsAsFactors = FALSE
)

# Create the forward (lead) dependent variable
# Stata: xtset country_id trend; F.e_migdpgro_5yr
# F. in Stata advances by one panel time unit (trend), which is one 5-year period
dat <- dat %>%
  arrange(country_id, trend) %>%
  group_by(country_id) %>%
  mutate(F_e_migdpgro_5yr = lead(e_migdpgro_5yr, 1)) %>%
  ungroup()

# Create factor variables for fixed effects
```

```
dat$country_f <- factor(dat$country_id)
dat$year_f <- factor(dat$year)
```

The dataset contains 9876 country-period observations for 424 countries over 1875–2010 (5-year intervals).

Note on variable names: The `.tab` export uses `e_polity2_ipo` and `e_miurbaniz_ipo` where the Stata do-file refers to `e_polity2` and `e_miurbaniz`. These are the same underlying variables.

3 Specification

The Stata command for Table 1, Model 2 is:

```
xtreg F.e_migdpgro_5yr upu_totalpo_ipo_ln_stock_1_5yr e_migdppcln_5yr
  e_mipopula_ipo_ln e_miurbaniz e_polity2 i.year, fe vce(cluster country_id) dfadj
```

The variables are:

Role	Variable	Description
Y	`F.e_migdpgro_5yr`	5-year GDP growth (1 period forward)
D	`upu_totalpo_ipo_ln_stock_1_5yr`	Log cumulative post offices stock (5-yr lag)
Z_1	`e_migdppcln_5yr`	Log GDP per capita (5-yr lag)
Z_2	`e_mipopula_ipo_ln`	Log population
Z_3	`e_miurbaniz_ipo`	Urbanization rate
Z_4	`e_polity2_ipo`	Polity2 democracy score
FE	`country_id` + `year`	Country and year fixed effects

4 Replication of Table 1

4.1 Model 1 (Parsimonious)

```
mod1 <- feols(
  F_e_migdpgro_5yr ~ upu_totalpo_ipo_ln_stock_1_5yr + e_migdppcln_5yr |
    country_f + year_f,
  data = dat, vcov = ~country_id
)
```

4.2 Model 2 (Full controls)

```
mod2 <- feols(
  F_e_migdpgro_5yr ~ upu_totalpo_ipo_ln_stock_1_5yr + e_migdppcln_5yr +
    e_mipopula_ipo_ln + e_miurbaniz_ipo + e_polity2_ipo |
    country_f + year_f,
  data = dat, vcov = ~country_id
)
```

```
comparison <- data.frame(
  Variable = c("Post offices (ln stock)", "GDP p.c. (ln)", "Population (ln)",
               "Urbanization", "Polity2"),
  Model_1_coef = c(
    round(coef(mod1)["upu_totalpo_ipo_ln_stock_1_5yr"], 4),
    round(coef(mod1)["e_migdppcln_5yr"], 4),
    NA, NA, NA
  ),
```

```
  Model_1_SE = c(
    round(sqrt(diag(vcov(mod1)))["upu_totalpo_ipo_ln_stock_1_5yr"], 4),
    round(sqrt(diag(vcov(mod1)))["e_migdppcln_5yr"], 4),
    NA, NA, NA
  ),
  Model_2_coef = c(
    round(coef(mod2)["upu_totalpo_ipo_ln_stock_1_5yr"], 4),
    round(coef(mod2)["e_migdppcln_5yr"], 4),
    round(coef(mod2)["e_mipopula_ipo_ln"], 4),
    round(coef(mod2)["e_miurbaniz_ipo"], 4),
    round(coef(mod2)["e_polity2_ipo"], 4)
  ),
  Model_2_SE = c(
    round(sqrt(diag(vcov(mod2)))["upu_totalpo_ipo_ln_stock_1_5yr"], 4),
    round(sqrt(diag(vcov(mod2)))["e_migdppcln_5yr"], 4),
    round(sqrt(diag(vcov(mod2)))["e_mipopula_ipo_ln"], 4),
    round(sqrt(diag(vcov(mod2)))["e_miurbaniz_ipo"], 4),
    round(sqrt(diag(vcov(mod2)))["e_polity2_ipo"], 4)
  )
)
kable(comparison, digits = 4,
      caption = "Replication of Table 1: Country panel TWFE models",
      col.names = c("Variable", "Coef (M1)", "SE (M1)", "Coef (M2)", "SE (M2)"))
```

Table 2: Replication of Table 1: Country panel TWFE models

	Variable	Coef (M1)	SE (M1)	Coef (M2)	SE (M2)
upu_totalpo_ipo_ln_stock_1_5yr	Post offices (ln stock)	0.0129	0.0042	0.0198	0.0054
e_migdppcln_5yr	GDP p.c. (ln)	-2.9849	0.4409	-4.0130	0.6189
e_mipopula_ipo_ln	Population (ln)	NA	NA	-1.9465	0.6311
e_miurbaniz_ipo	Urbanization	NA	NA	-3.1421	2.9731
e_polity2_ipo	Polity2	NA	NA	0.0341	0.0210

Model 1 uses 1835 observations across 153 countries. Model 2 uses 1540 observations across 147 countries. The treatment coefficient on post offices is $\hat{\beta} = 0.0198$ ($p = 4 \times 10^{-4}$) in Model 2, statistically significant and positive. Standard errors may differ slightly from the published values due to minor differences in Stata's `xtreg, fe dfadj` degrees-of-freedom adjustment and `fixest`'s default small-sample correction.

5 IVB Decomposition

5.1 Setup

We define:

- **Treatment** (D): `upu_totalpo_ipo_ln_stock_1_5yr` (lagged log cumulative post offices)
- **Outcome** (Y): `F_e_migdpgro_5yr` (5-year GDP growth, forward)
- **Candidate colliders** (z): each of the four controls, one at a time
- **Fixed effects**: country + year

For each candidate collider z_j, the "short" model excludes z_j (but retains the other controls), and the "long" model includes all controls. The IVB identity $\hat{\beta}^{\text{long}} - \hat{\beta}^{\text{short}_j} = -\hat{\theta}_j \cdot \hat{\pi}_j$ is verified numerically.

5.2 Decomposition

```r
d_var <- "upu_totalpo_ipo_ln_stock_1_5yr"
all_controls <- c("e_migdppcln_5yr", "e_mipopula_ipo_ln",
                  "e_miurbaniz_ipo", "e_polity2_ipo")
control_labels <- c("Log GDP p.c.", "Log population",
                    "Urbanization", "Polity2")

ivb_results <- lapply(seq_along(all_controls), function(j) {
  z_var <- all_controls[j]
  w_vars <- setdiff(all_controls, z_var)
  compute_ivb_multi(
    data = dat, y = "F_e_migdpgro_5yr", d_vars = d_var,
    z = z_var, w = w_vars,
    fe = c("country_f", "year_f"), vcov = "iid"
  )
})
names(ivb_results) <- all_controls

tab_ivb <- do.call(rbind, lapply(seq_along(all_controls), function(j) {
  r <- ivb_results[[j]]$results
  data.frame(
    Collider = control_labels[j],
    beta_short = r$beta_short,
    beta_long = r$beta_long,
    theta = r$theta,
    pi = r$pi,
    IVB = r$ivb_formula,
    pct_beta = round(100 * r$ivb_formula / r$beta_long, 2),
    check = r$diff_check,
    stringsAsFactors = FALSE
  )
}))
```

```r
kable(tab_ivb, digits = 6, row.names = FALSE,
      caption = "IVB decomposition for each candidate collider (Table 1, Model 2)",
      col.names = c("Candidate collider $z_j$",
                    "$\\hat\\beta_{\\text{short}_j}$",
                    "$\\hat\\beta_{\\text{long}}$",
                    "$\\hat\\theta_j$",
                    "$\\hat\\pi_j$",
                    "IVB $= -\\hat\\theta_j \\hat\\pi_j$",
                    "IVB / $\\hat\\beta_{\\text{long}}$ (\\%)",
                    "Check"))
```

Table 3: IVB decomposition for each candidate collider (Table 1, Model 2)

Candidate collider z_j	$\hat\beta_{\text{short}_j}$	$\hat\beta_{\text{long}}$	$\hat\theta_j$	$\hat\pi_j$	IVB $= -\hat\theta_j\hat\pi_j$	IVB / $\hat\beta_{\text{long}}$ (%)	Check
Log GDP p.c.	0.008333	0.019817	-4.012961	0.002862	0.011485	57.95	0
Log population	0.020619	0.019817	-1.946508	0.000412	-0.000802	-4.05	0

4

Candidate collider z_j	$\hat{\beta}_{\text{short}_j}$	$\hat{\beta}_{\text{long}}$	$\hat{\theta}_j$	$\hat{\pi}_j$	IVB $= -\hat{\theta}_j\hat{\pi}_j$	IVB / $\hat{\beta}_{\text{long}}$ (%)	Check
Urbanization	0.019219	0.019817	- 3.142092	0.000191	0.000599	3.02	0
Polity2	0.019065	0.019817	0.034098	- 0.022055	0.000752	3.79	0

The identity holds exactly (check ≈ 0) in all cases, confirming the numerical validity of the decomposition.

5.3 Interpretation of IVB components

```r
tab_summary <- data.frame(
  Control = control_labels,
  theta = round(tab_ivb$theta, 4),
  pi = round(tab_ivb$pi, 6),
  IVB = round(tab_ivb$IVB, 6),
  pct = round(tab_ivb$pct_beta, 1),
  Direction = ifelse(tab_ivb$IVB > 0, "Inflates", "Deflates")
)

kable(tab_summary, digits = 4, row.names = FALSE,
      caption = "Summary of IVB components",
      col.names = c("Control", "$\\hat\\theta_j$ (effect on $Y$)",
                    "$\\hat\\pi_j$ (effect of $D$ on $Z_j$)",
                    "IVB", "IVB/$\\hat{\\beta}_{\\text{long}}$ (\\%)",
                    "Direction"))
```

Table 4: Summary of IVB components

Control	$\hat{\theta}_j$ (effect on Y)	$\hat{\pi}_j$ (effect of D on Z_j)	IVB	IVB/$\hat{\beta}_{\text{long}}$ (%)	Direction
Log GDP p.c.	-4.0130	0.0029	0.0115	58.0	Inflates
Log population	-1.9465	-0.0004	- 0.0008	-4.0	Deflates
Urbanization	-3.1421	0.0002	0.0006	3.0	Inflates
Polity2	0.0341	-0.0221	0.0008	3.8	Inflates

5.4 Key findings

```r
total_ivb <- sum(tab_ivb$IVB)
beta_long <- tab_ivb$beta_long[1]

# Short model with NO controls (only FE)
mod_no_controls <- feols(
  F_e_migdpgro_5yr ~ upu_totalpo_ipo_ln_stock_1_5yr |
    country_f + year_f,
  data = dat[complete.cases(dat[, c("F_e_migdpgro_5yr", d_var,
    all_controls, "country_f", "year_f")]), ],
  vcov = "iid"
)
beta_no_controls <- coef(mod_no_controls)[d_var]
```

- $\hat{\beta}_{\mathbf{long}}$ (with all controls) $= 0.0198$
- $\hat{\beta}_{\mathbf{no\ controls}}$ (only FE, same sample) $= 0.0067$
- **Total IVB** (sum across all controls) $= 0.012034$ (60.7% of $\hat{\beta}_{\mathrm{long}}$)

The IVB decomposition reveals that **log GDP per capita** is by far the most influential control, contributing an IVB of 0.0115 — approximately 58% of $\hat{\beta}_{\mathrm{long}}$. Including GDP per capita as a control substantially *increases* the estimated treatment effect of post offices (from 0.0083 to 0.0198). This is because:

1. $\hat{\theta}_1 < 0$: Higher GDP per capita is associated with *lower* subsequent GDP growth (conditional convergence), and
2. $\hat{\pi}_1 > 0$: More post offices predict *higher* GDP per capita.

The product $-\hat{\theta}_1 \cdot \hat{\pi}_1 > 0$, inflating the treatment coefficient when GDP per capita is included.

The remaining three controls (population, urbanization, Polity2) contribute relatively small IVBs, each less than 4% of $\hat{\beta}_{\mathrm{long}}$ in absolute terms.

6 Literature Support for Collider Status

The IVB formula diagnoses **how much** each control shifts the treatment estimate, but **whether** a control is a collider depends on substantive knowledge: does postal infrastructure cause the control, and does future GDP growth also cause the control (or are they both driven by a common cause)?

6.1 Post offices $\to$ GDP per capita (collider channel for Z_1)

There is strong evidence that postal infrastructure affects economic development, which in turn determines GDP per capita levels:

- **Rogowski et al. (2022)** themselves argue that post offices facilitated economic development through information dissemination, market integration, and institutional capacity building. The positive $\hat{\pi}_1$ (post offices predict higher GDP p.c.) is precisely the mechanism the paper studies.
- **Acemoglu, Johnson & Robinson (2001, AER):** Institutional infrastructure (of which postal systems are a component) shapes long-run income differences across countries.

At the same time, GDP growth (the outcome) mechanically determines future GDP per capita levels. If GDP growth at $t+1$ raises GDP per capita at $t+1$, and the lagged GDP per capita at t is correlated with growth at $t+1$ through convergence dynamics, then GDP per capita is on the causal pathway between postal infrastructure and GDP growth.

Implication: GDP per capita is a plausible collider. Controlling for it absorbs part of the effect of post offices that operates through raising GDP levels, creating the large positive IVB observed. However, GDP per capita also serves as a *confounder* (richer countries may both build more post offices and have different growth trajectories), which justifies its inclusion. The direction of net bias depends on which channel dominates.

6.2 Post offices $\to$ Population (Z_2)

- **Infrastructure and migration:** Public infrastructure can attract migration and population growth. However, causality likely runs primarily in the reverse direction: larger populations demand more post offices. The very small IVB ($< 1\%$ of $\hat{\beta}$) suggests this control is empirically innocuous.

6.3 Post offices $\to$ Urbanization (Z_3)

- **Infrastructure and urbanization:** Postal networks may facilitate urbanization by connecting rural areas to urban markets, reducing information costs for migration decisions. However, the IVB is very small (about 3% of $\hat{\beta}$) and the collider concern is modest.

6.4 Post offices → Polity2 (Z_4)

- **Infrastructure and democratization:** There is some evidence that communication infrastructure facilitates political participation and democratization:
  - **Prat & Stromberg (2013, ARE):** Media and communication infrastructure affect political accountability and democratic participation.
  - **Rogowski et al.** note that postal systems were historically linked to state capacity and political development.
- The IVB from Polity2 is small (about 4% of $\hat{\beta}$), suggesting limited empirical concern.

7 Summary

```r
summary_all <- data.frame(
  Control = control_labels,
  `Y causes Z?` = c(
    "Yes (mechanical: growth -> GDP level)",
    "Weak/reverse",
    "Weak",
    "Possible"
  ),
  `D causes Z?` = c(
    "Yes (core mechanism)",
    "Weak/reverse",
    "Possible",
    "Possible"
  ),
  `Collider?` = c(
    "Plausible (also confounder)",
    "Unlikely",
    "Unlikely",
    "Unlikely"
  ),
  IVB = round(tab_ivb$IVB, 4),
  `IVB pct` = round(tab_ivb$pct_beta, 1),
  check.names = FALSE, stringsAsFactors = FALSE
)
kable(summary_all,
      caption = "Summary: Literature evidence and IVB magnitudes",
      col.names = c("Control", "$Y \\to Z$?", "$D \\to Z$?",
                    "Collider?", "IVB", "IVB/$\\hat{\\beta}$ (\\%)"))
```

Table 5: Summary: Literature evidence and IVB magnitudes

Control	$Y \to Z$?	$D \to Z$?	Collider?	IVB	IVB/$\hat{\beta}$ (%)
Log GDP p.c.	Yes (mechanical: growth -> GDP level)	Yes (core mechanism)	Plausible (also confounder)	0.0115	58.0
Log population	Weak/reverse	Weak/reverse	Unlikely	-0.0008	-4.0
Urbanization	Weak	Possible	Unlikely	0.0006	3.0
Polity2	Possible	Possible	Unlikely	0.0008	3.8

The main finding is that **log GDP per capita** dominates the IVB decomposition, accounting for 58% of the

long-model treatment coefficient. Including GDP per capita *inflates* the estimated effect of post offices from 0.008 (short model) to 0.02 (long model). This is a consequence of conditional convergence: controlling for GDP per capita level absorbs a negative confound (richer countries grow more slowly) while simultaneously opening a collider pathway (post offices raise GDP per capita, which is negatively associated with growth given convergence).

The fact that $\hat{\beta}_{\text{long}} > \hat{\beta}_{\text{short}}$ when GDP per capita is included is a well-known pattern in growth regressions: the effect of any growth determinant typically appears larger once one controls for initial income, because the convergence effect is partialled out. Whether this represents bias (collider) or deconfounding depends on whether GDP per capita at time t is purely pre-determined (confounder) or also reflects the causal pathway from post offices to growth (collider/mediator). In a dynamic panel setting with 5-year intervals, both channels are likely operative.

The remaining controls contribute negligible IVB, suggesting the model's sensitivity to control specification is driven almost entirely by the convergence control.

8 References

- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5), 1369–1401.
- Prat, A. & Stromberg, D. (2013). The Political Economy of Mass Media. In *Advances in Economics and Econometrics* (10th World Congress), Vol. 2. Cambridge University Press.
- Rogowski, J. C., Gerring, J., Maguire, M., & Cojocaru, L. (2022). Public Infrastructure and Economic Development: Evidence from Postal Systems. *American Journal of Political Science*, 66(4), 885–901.