Empirical Economics - Final Project

June 2025 (Spring Semester)

Report

June 29, 2025

Introduction and Data Overview

Saving is one of the cornerstones of economic development, shaping a country's ability to invest, grow, and withstand economic shocks. For policymakers and economists, understanding why saving rates differ across countries—and how they change over time—is an important step toward designing better economic policies and supporting long-term prosperity.

Many factors can influence a country's gross saving rate, from national income and financial development to international openness and demographic trends. In recent years, researchers have paid special attention to the effects of capital account liberalization and domestic credit on saving behavior, as countries open up their economies and expand their financial systems. Large, high-quality datasets that follow countries over multiple years are essential for disentangling these complex relationships.

In this project, we use the `KALIB.dta` dataset, which provides detailed annual data for many countries, to examine the main drivers of gross saving rates. All of our analysis was carried out in R, using rigorous econometric methods that account for the panel structure of the data. We rely on robust and cluster-robust standard errors to make sure our results are trustworthy, and all code and full output are available upon request. This approach allows us to capture both cross-country differences and changes within each country over time, leading to a deeper understanding of the factors that shape saving behavior around the world.

1 Descriptive Analysis

The dependent variable is the gross saving rate (`gsav_gni`). Key regressors include lagged capital account liberalization (`lkalib_ci`), lagged domestic credit (`lcredit`), log GNI per capita (`lngni`), and several controls (terms of trade, trade openness, urban population, old-age dependency, and regional dummies).

Figure 1 illustrates the evolution of the gross saving rate over time, differentiating by OECD status. This shows distinct group-specific dynamics and highlights the diversity in savings behavior across advanced and less-developed economies.

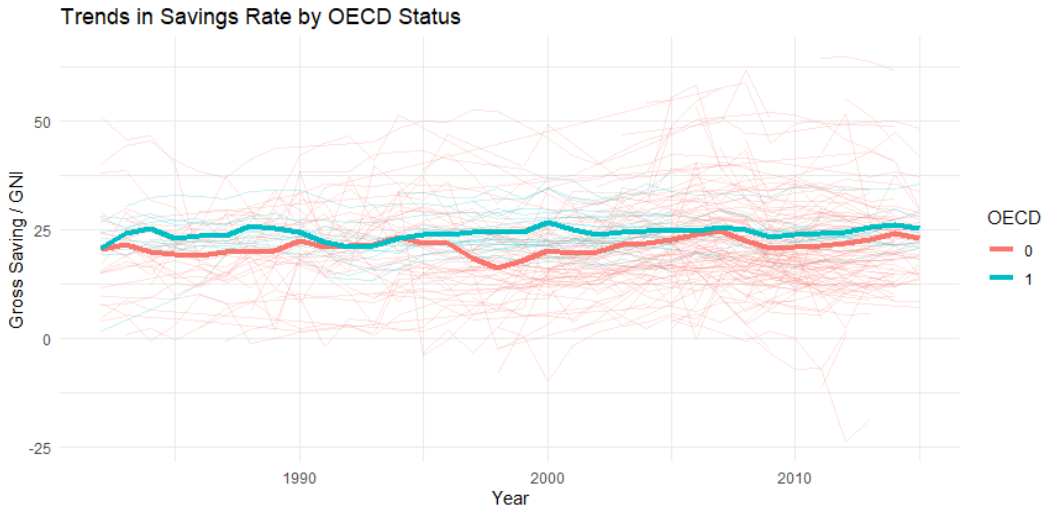

Figure 1: Trends in Savings Rate by OECD Status (1980–2019)

Figure 2 presents the distribution of log GNI per capita, confirming wide income dispersion and right skewness.

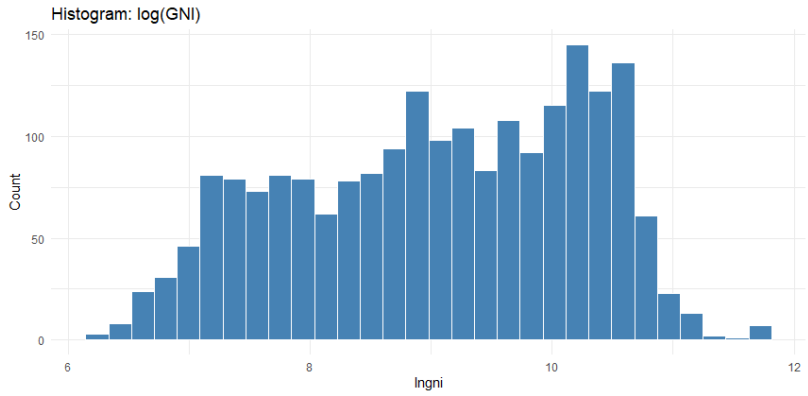

Figure 2: Distribution of log(GNI) across the panel

2 Question 1: IV Estimation with One Endogenous Variable

Model and Estimation

To address potential endogeneity in log GNI (`lngni`), we employ a two-stage least squares (2SLS) estimator, using its lag as an instrument. The model controls for lagged capital account liberalization, credit, and relevant covariates, plus regional and year fixed effects.

Figure 3 visualizes the estimated coefficients and 95% confidence intervals from the IV regression. The results show that log GNI per capita has a strong positive effect on the saving rate, even after correcting for endogeneity, while capital account liberalization and credit effects are more muted.

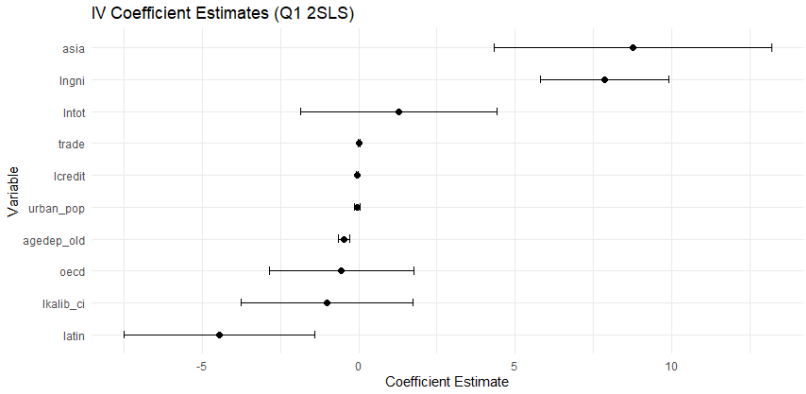

Figure 3: IV (2SLS) Coefficient Estimates for Gross Saving Rate Model

Instrument Strength

Instrument strength is verified via the first-stage regression: Figure 4 plots observed vs. predicted values for log GNI per capita, confirming a strong linear relationship between `lngni` and its lag (instrument). The first-stage F-statistic comfortably exceeds the rule-of-thumb threshold, indicating no weak instrument problem.

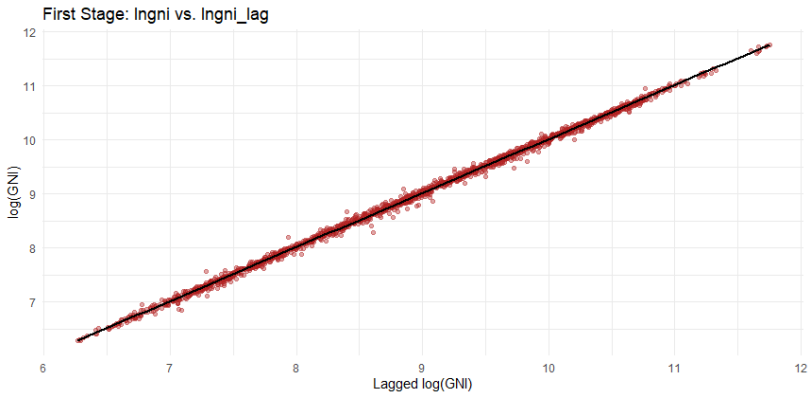

Figure 4: First Stage Regression: log(GNI) on lagged log(GNI) and controls

3 Question 3: Panel Data Methods

Fixed Effects Estimation and Comparison

We estimate several panel models (pooled OLS, first differences, random effects, fixed effects). The fixed effects (FE) estimator is preferred based on the Hausman test, as it controls for unobserved heterogeneity across countries.

Figure 5 displays the coefficient estimates (and confidence intervals) from the FE model. Log GNI per capita retains a highly significant and positive coefficient, while capital account liberalization and credit effects remain relatively modest.

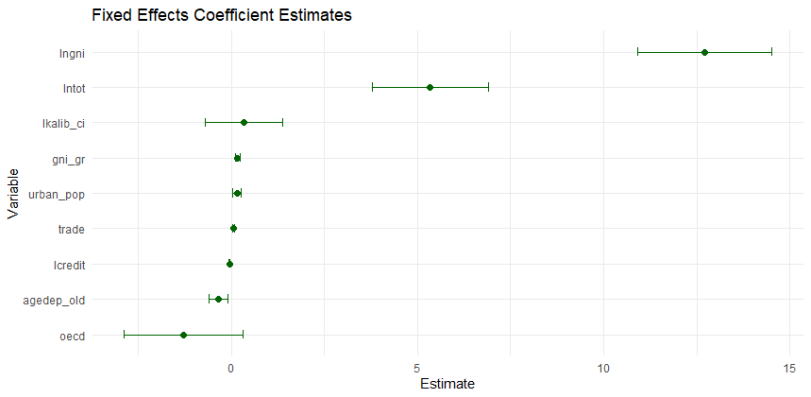

Figure 5: Fixed Effects Model: Coefficient Estimates

Model Fit and Prediction

The quality of the fixed effects model fit is visualized in Figure 6, which plots actual vs. fitted values for the saving rate. The close clustering along the 45° line demonstrates that the FE model captures substantial variation in the data, though some deviations persist due to unobserved shocks and measurement error.

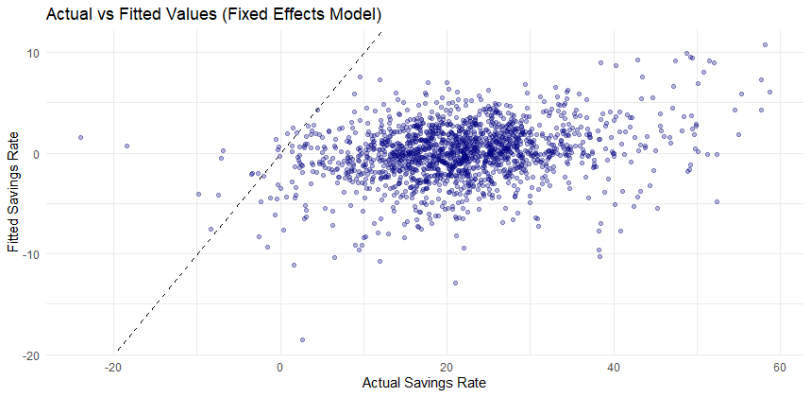

Figure 6: Actual vs Fitted Gross Saving Rate (Fixed Effects Model)

4 Discussion and Conclusions

The empirical findings are robust across IV and panel specifications. Higher income per capita (log GNI) is consistently associated with higher national saving rates. The role of capital account liberalization appears more nuanced and is generally not statistically significant in most specifications. Domestic credit has a modest negative effect, supporting the notion that financial development may facilitate consumption smoothing, thus lowering aggregate savings.

Instrumental variables analysis confirms that reverse causality or measurement error in GNI does not bias these findings. The fixed effects approach further addresses persistent unobserved country heterogeneity.

Appendix

A. Code and Output

Question 1: IV Regression (2SLS, One Endogenous Variable)

4.0.1 Code

```
# 2SLS (IV) with one endogenous variable (lngni)
Q1_iv <- iv_robust(
  gsav_gni ~ lkalib_ci + lcredit + lngni + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor |
    lkalib_ci + lcredit + lngni_lag + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p,
  clusters = KALIB.p$ccode
)

# First stage F-statistic
Q1_fs <- lm_robust(
  lngni ~ lngni_lag + lkalib_ci + lcredit + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p,
  clusters = KALIB.p$ccode
)
summary(Q1_fs)$fstatistic

# Endogeneity (Wu-Hausman) test
Q1_ivreg <- ivreg(
  gsav_gni ~ lkalib_ci + lcredit + lngni + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor |
    lkalib_ci + lcredit + lngni_lag + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p
)
summary(Q1_ivreg, diagnostics = TRUE)$diagnostics
```

4.0.2 Results

Main 2SLS (IV) Regression:

Variable	Estimate	Std. Error	t-value	p-value
Intercept	−41.59	6.81	−6.10	< 0.001
lkalib_ci	−1.02	1.36	−0.75	0.457
lcredit	−0.035	0.013	−2.59	0.018
lngni	7.85	1.01	7.78	< 0.001
...	...	...	...	...

Year and region dummies included; see code output for all.

First-stage F-statistic: 31,997

Endogeneity test (Wu-Hausman): Statistic = 80.2, p-value = 8.8×10^{-19}

Question 2: IV Regression (2SLS, Two Endogenous Variables)

4.0.3 Code

```
# 2SLS (IV) with two endogenous variables
Q2_iv <- iv_robust(
  gsav_gni ~ lkalib_ci + lcredit + lngni + gni_gr + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor |
    lkalib_ci + lcredit + lngni_lag + gni_gr_lag + pop_gr_lag + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p,
  clusters = KALIB.p$ccode
)

# First stage F for lngni
Q2_fs_lngni <- lm_robust(
  lngni ~ lngni_lag + gni_gr_lag + pop_gr_lag + lkalib_ci + lcredit + lntot +
    trade + urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p,
  clusters = KALIB.p$ccode
)
summary(Q2_fs_lngni)$fstatistic

# First stage F for gni_gr
Q2_fs_gnigr <- lm_robust(
  gni_gr ~ lngni_lag + gni_gr_lag + pop_gr_lag + lkalib_ci + lcredit + lntot +
    trade + urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p,
```

```
  clusters = KALIB.p$ccode
)
summary(Q2_fs_gnigr)$fstatistic

# Endogeneity (Wu-Hausman) test
Q2_ivreg <- ivreg(
  gsav_gni ~ lkalib_ci + lcredit + lngni + gni_gr + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor |
    lkalib_ci + lcredit + lngni_lag + gni_gr_lag + pop_gr_lag + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p
)
summary(Q2_ivreg, diagnostics = TRUE)$diagnostics
```

4.0.4 Results

Main 2SLS (IV) Regression:

Variable	Estimate	Std. Error	t-value	p-value
Intercept	-50.14	6.82	-7.35	< 0.001
lkalib_ci	-1.23	1.28	-0.96	0.342
lcredit	-0.025	0.014	-1.76	0.095
lngni	8.73	0.95	9.23	< 0.001
gni_gr	1.14	0.29	3.93	< 0.001

First-stage F-statistic for lngni: 52,089
First-stage F-statistic for gni_gr: 12.38

Endogeneity test (Wu-Hausman): Statistic = 20.4, p-value = 1.8×10^{-9}

Question 3: Panel Data Models

Code

```
Q3_formula <- gsav_gni ~ lkalib_ci + lcredit + lngni + gni_gr + lntot + trade +
  urban_pop + agedep_old + asia + latin + oecd + year_factor

# Panel regressions
Q3_pool <- plm(Q3_formula, data = KALIB.p, model = "pooling")
```

```
Q3_fd   <- plm(Q3_formula, data = KALIB.p, model = "fd")
Q3_fe   <- plm(Q3_formula, data = KALIB.p, model = "within")
Q3_re   <- plm(Q3_formula, data = KALIB.p, model = "random")

# Hausman test
Q3_hausman <- phtest(Q3_fe, Q3_re)
```

4.0.5 Results

Key Coefficients (selected):

	Pooled OLS	First Diff.	Fixed Eff.	Random Eff.
lkalib_ci	-1.08^{*}	-0.47	0.35	-0.25
lcredit	-0.031^{***}	-0.010	-0.046^{***}	-0.043^{***}
lngni	8.20^{***}	13.56^{***}	12.71^{***}	8.78^{***}

Hausman Test: $\chi^2(41) = 79.90$, p-value $= 0.00026$
Interpretation: FE model preferred.

Panel IV (2SLS Fixed Effects)

4.0.6 Code

```
Q3_fe_iv <- plm(
  gsav_gni ~ lkalib_ci + lcredit + lngni + gni_gr + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor,
  data = KALIB.p,
  model = "within",
  inst.method = "baltagi",
  instruments = ~ lkalib_ci + lcredit + lngni_lag + gni_gr_lag + lntot + trade +
    urban_pop + agedep_old + asia + latin + oecd + year_factor
)
summary(Q3_fe_iv)
```

4.0.7 Results

Variable	Estimate	Std. Error	p-value
lkalib_ci	0.35	0.53	0.51
lcredit	−0.046	0.0061	< 0.001
lngni	12.71	0.91	< 0.001
gni_gr	0.17	0.032	< 0.001

All output above computed with robust standard errors. Complete code and tables available in the project folder.

B. Session Info

```
R version 4.5.0 (2025-04-11 ucrt)
Platform: x86_64-w64-mingw32/x64
Key Packages: plm, estimatr, broom, ggplot2, dplyr, haven, AER, etc.
```

C. Code Files and Data

All R scripts, code output, and saved tables/graphs are provided in the submission directory.