ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
SCHOOL OF ECONOMIC SCIENCES
ΤΜΗΜΑ ΔΙΕΘΝΩΝ & ΕΥΡΩΠΑΪΚΩΝ ΟΙΚΟΝΟΜΙΚΩΝ ΣΠΟΥΔΩΝ
DEPARTMENT OF INTERNATIONAL & EUROPEAN ECONOMIC STUDIES

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203197 / Fax: 210 8214122
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203197 / Fax: (+30) 210 8214122
E-mail: pkonstantinou@aueb.gr / www.aueb.gr

Empirical Economics
Final Project

June 2025 (Spring Semester 2025) **Date of Posting**: **06/06/2025**

General Instructions

The answers to the questions should be submitted via the asynchronous eClass system under the link Assignments. All submissions should take the form of compressed files «.zip» (https://www.7-zip.org/) and you should NOT select the option "Uncompress zip file on the server".[1] Each submitted answer should include a text document with all your answers (tables, equations when necessary and text/discussion where required) and one or more files of all the calculations performed in **R**. In particular, you should have a file (or files) in R (e.g. PKProject.R from RStudio) which will contain all the steps you performed along the way: «loading» data, creating new variables and data.frames when necessary, estimating regressions etc. In all your estimations you should employ cluster robust standard errors. These can be easily introduced in the `estimatr` package or the `plm` package.[2] You may build upon the file panel data examples I have uploaded in the [codes] folder and consult the solution to Assignment 3. In addition, I would advise you to define the dataset as a panel data frame from the beginning, regardless of whether or not you use panel data methods for the estimation.

Submission Deadline: **Thursday, July 3, 2025 (03/07/2025)**

The questions will be answered using KALIB.dta dataset available in eClass.[3] A short description of the dataset is included with the data (KALIB_description.txt). The dataset includes data on the saving rate of various countries around the world, and various key determinants of these saving rates. The dependent variable of interest is the gross saving rate `gsav_gni`, defined as the gross national saving as a percentage of Gross National Income (GNI). The key exogenous control variables include (*i*) the log of the terms of trade (`lntot`) and international trade (`trade`), (ii) demographics such as `urban_pop` and

[1] This is necessary as eClass accepts Word and pdf file but not files with "funny" extensions (like Assignment.R).

[2] For detailed instructions about using the `estimatr` package you should consult the dedicated web page https://cran.r-project.org/web/packages/estimatr/index.html, and focus especially on the routines `lm_robust()` & `iv_robust()`. For the `plm` package consult https://cran.r-project.org/web/packages/plm/ .

[3] The data are provided in STATA (.dta) format and in Excel. The former format of the dataset can easily be imported in R using the package `haven`:

```
library(haven)
KALIB <- read_dta("…/KALIB.dta")
```

Otherwise just use the relevant excel file.

agedep_old, (iii) degree of financial development in year $t-1$ (lcredit), and (iv) capital account liberalization index in year $t-1$ (lkalib_ci).[4] The endogenous variables are the log of GNI per person (lngni) and growth rate of GNI (gni_gr).

We want to estimate relationships of the form

$$s_{i,t} = \gamma_1 KALIB_{i,t-1} + \gamma_2 credit_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{i,t} + \boldsymbol{\xi}'\mathbf{w}_{i,t} + reg_i + a_i + \lambda_t + \varepsilon_{i,t} \quad (1)$$

where $s_{i,t}$ is the saving rate in country i during year t, $KALIB_{i,t-1}$ is a indicator of (relatively) free capital movements from/to country i during year $t-1$, $credit_{i,t-1}$ is an indicator of the development of the domestic financial system, $\mathbf{w}_{i,t}$ contains the exogenous variables/controls (lntot, trade, urban_pop, agedep_old) and $\boldsymbol{x}_{i,t}$ contains the endogenous variables (lngni, gni_gr). The a_i control for unobserved heterogeneity, the λ_t denote time dummies while reg_i are dummies about the region in which country i belongs. We are mostly interested in the parameters γ_1 and γ_2, but also those in the vectors $\boldsymbol{\xi}$ and $\boldsymbol{\beta}$ are important.

Instrumental Variables Estimation and Endogeneity

1. Here the basic regression is (1) where we set $a_i = 0$ (we forget time invariant heterogeneity) and where the only endogenous variable is lngni – all the other variables and dummies are included in the model (gni_gr is not included in the regression, only lngni).
    a. Estimate the model using 2SLS where the instrument for lngni is its time lag.[5]
    b. Is the instrumental variable relevant?
    c. Perform a test of whether the endogenous variable lngni is exogenous or not.
    d. What do your estimates show about the effects of the explanatory variables show?
2. Generalizing the above analysis, we set again $a_i = 0$ but now we have two endogenous variables, lngni and gni_gr (now both lngni and gni_gr are included in the regression).
    a. Estimate the model using 2SLS where the instruments are now the time lags of lngni and gni_gr as well as the time lag of the population growth rate (pop_gr).
    b. Are the instrumental variables relevant?
    c. Can the endogenous variables be treated as exogenous? Perform a test to assess this.

[4] See Loyaza, Hebbel-Schmidt, & Serven (2000) for the general philosophy behind these.
[5] In the context of the plm package this can be constructed as: `KALIB.p$lngnil <- plm::lag(KALIB.p$lngni, k = 1)`.

d. What do your estimates show about the effects of the explanatory variables show? How do these differ relative to what you found in question 1?

Panel Estimation Methods

3. We return to equation (1) again where the endogenous variables are again `lngni` and `gni_gr`, but for the time being treat them as exogenous.
    a. Estimate the model using pooled OLS, first differences, fixed effects and random effects and compare your results across models.
    b. If you had to choose between a model with fixed effects and a model with random effects which one would you prefer? Present a test to support your claim.
    c. How much would your results change if you change `lcredit` in your model with the lagged value of `FD`?
    d. Go back to the models you estimated in 3(a) and focus on the fixed effects and random effects models. Estimate the models using 2SLS for panels.[6]

[6] If you check the help of the `plm` package (`?plm`) this is a straightforward extension of what you did in questions 1 and 2 above. Use the default options.