

Empirical Project 2  
Due: August 11, 2021

Note: the relevant files for this project are on ELMS in the “Files->Empirical Project 2” folder

Suppose you are asked to estimate the determinants of smoking cigarettes among youth in developing countries. In this project you are asked to do a mini research analysis where you will decide on the suitable model, collect and construct data, estimate the model and conduct inference on the estimated probability model of smoking. We will focus on African region and use Global Youth Tobacco Survey (GYTS) developed by the Centers for Disease Control and Prevention (CDC) to track tobacco use of young people. The questionnaire is the same across countries making it suitable for the cross-country analysis.

The following steps will guide the development of this project:

- (a) Read the article “DOES ADVERTISING MATTER? ESTIMATING THE IMPACT OF CIGARETTE ADVERTISING ON SMOKING AMONG YOUTH IN DEVELOPING COUNTRIES” by Kostova and Blecher. In the introduction to your presentation/briefing paper briefly discuss:
  - What is the research objective of the paper?
  - What are the empirical issues based on the existing research on this topic that are pointed out by authors?
  - Summarize the methodology to approach the research objective
  - Summarize the empirical results of the paper. Are the results consistent with the hypothesis of the paper?
- (b) Collect the data:
  - Global Youth Tobacco Survey (GYTS) is free publicly accessible data: <https://nccd.cdc.gov/GTSSDataSurveyResources/Ancillary/DataReports.aspx?CAID=2> and select Youth from the drop down menu (Population (Survey))
  - Select the WHO Region: African (AFR);
  - Select a country: for this project pick any two African countries in a chosen year (you will have to separately download survey data for each country) Please, make sure that for each country you will pick the chosen survey year is the same( for example Ghana national for 2009, and Congo national for 2009);
  - Each survey comes in .zip file that contains data in .mdb format and data dictionary (survey questionnaire) in .pdf format. In order to open .mdb file you will need Windows PC. In Excel, navigate to the folder you have stored the .mdb file. Select all files; and open the .mdb file. The pop-up screen will

appear “Microsoft Excel Security Notice” – click Enable; Import data pop-up screen will appear – click OK. Save this file as .xlsx. You can then import this file into STATA.

(c) Prepare the data:

- The data we are working with in this project has a survey structure that has primary sampling unit variable (PSU); sampling weight (FinalWgt) and strata (Stratum). In STATA we use the following command to declare data as survey: `svyset [psu] [pweight], strata( )`. Use STATA help `svyset` to read how to use this command to set the data for this project.
- Each column in the data corresponds to a question in the survey found in the accompanying .pdf file. For example: CR1 corresponds to Question 1 – How old are you. Note that each question is a categorical variable. Please model these, accordingly, noting the base category.

(d) Design the Model:

- Using the design in the leading article, we will estimate the probability model of smoking assuming an individual from the survey smokes at least one cigarette.
- Select questions of interest from the questionnaire as your determinants of smoking probability. (for comparison purposes, these determinants should be the same for each chosen country)

(e) Estimate the model:

- Since we are estimating using survey data, use `svy:` (e.g. `svy: probit y x`) prefix before running the probability models in STATA. For robustness you can run both probit and logit models. What, if any, differences do you notice? Which model is preferable?
- Ensure the models are robust to heteroskedasticity.
- Use margins command to obtain probabilities for each categorical variable in your model.

(f) Policy Implications:

- Summarize policy implications of the leading study.
- How does your own analysis differ for the two countries of your choice?
- Which factors lead to higher probability of smoking for the youth in your analysis?
- What are some policy recommendations that you can provide?