

Problem Set 5

Due on October 9, 2023, 10am

To successfully complete this problem set, please follow these steps:

1. Download the contents of the assignment Dropbox link in a folder (directory) on your computer designated for this problem set only. Follow the directory structure, e.g., putting the data folder containing the datasets as a subdirectory of the project directory.
2. Insert your answers in the yellow boxes using Microsoft Word, and prepare a single .R script for what you produce. Save the word document as a PDF.
3. Please submit the PDF to the designated PS-XX: pdf link and your R Script to the PS-XX: R link.

(1) Your name:

Nikhilla Bhuvana Sundar

(2) Group members, if any:

(3) Compliance with the Academic Code on problem set¹ (sign with an X below)

X

¹ You may use the same code from classmates, Ed Discussion Board, instructors, and generative AI. However, you must hand in your own unique written work and code in all cases. Any wholesale copy/paste of another's work is plagiarism. In other words, you can work with your classmate(s), sitting side-by-side and going through the problem set question-by-question, or use generative AI to provide potential code for you, but you must each type your own answers and your own code.

This problem set works with two datasets. Autograder will have the following packages loaded:

```
tidyverse
haven
modelsummary
fixest
```

Problem 1

In this problem, we will conduct a simple version of the analysis in the following article:

Edward Miguel, Shanker Satyanath, and Ernest Sergenti (2004). “Economic Shocks and Civil Conflict: An Instrumental Variables Approach”, *Journal of Political Economy*.

1.1

The Stata dataset `miguel_africa.dta` is the author’s dataset described in the paper. Each observation represents a country-year combination, and includes the following variables.

The paper’s authors are interested in the degree to which conflict is produced by economic shocks. They use change in rainfall from the previous year to instrument for economic shocks.

Variable name	Description
<code>year</code>	Year of measurement
<code>country_name</code> , <code>country_code</code> , <code>ccode</code>	Country (name, abbreviation, numerical code)
<code>any_prio</code>	A binary variable for the occurrence of any conflict fitting PRIO’s definition (25 battle deaths or more)
<code>polity2l</code>	The polity score of the country in that year, logged. This is an expert-coded score for how democratic a country is in that year. Ranges from -10 to 10 (most Democratic)
<code>GPCP_g</code>	Growth of Rainfall measured by the Global Precipitation Climatology Project (GPCP) at time t . See the paper p.733 for details of how growth (ΔR_t) is calculated.
<code>GPCP_g_l</code>	The growth in rainfall in the prior year, i.e. lagged at time $t - 1$.
<code>gdp_g</code>	The growth in GDP at year t , calculated the same way as rainfall.
<code>gdp_g_l</code>	The growth in GDP at year $t - 1$, calculated the same way as rainfall.

1.1

Create a single regression table with `modelsummary` with two regressions:

1. A simpler version of the author's instrumental variables regression, where:
 - Use the *lagged* economic growth (`gdp_g_l`) measure as the treatment measure.
 - Use the lagged rainfall growth (`GPCP_g_l`) to instrument for economic growth.
 - Use `any_prio` (whether or not each country year reported a conflict with more than 25 battle deaths a year) as the outcome.
 - Use the country-year's polity score (`polity2l`) as a control variable.
2. A naive OLS regression with the same outcome, treatment, and same control variable

As in Table 4 of Miguel et al., the treatment variable of interest should share the same row of the `modelsummary` table. You will need to use the `coef_map` argument in `modelsummary` to unify coefficient names on the treatment from the IV and OLS (See appendix).

	IV	OLS
Intercept	0.257	0.266
	(0.022)	(0.019)
Lagged Economic Growth	-2.468	-0.081
	(1.931)	(0.225)
Democracy (Polity IV)	0.001	0.000
	(0.003)	(0.003)
Num.Obs.	743	743

1.2

Interpret the coefficient estimate on the treatment variable substantively, in the context of this example. Your answer should follow the format from the class on units -- be careful of the units of each variable here.

OLS Model: As economic growth increases by 1 percentage point, incidence of conflicts reduces by approximately 0.08 percentage points.

IV Model: As economic growth increases by 1 percentage point, incidence of conflicts reduces by 2.5 percentage points approximately

Problem 2

We will start looking at the following article for our examination of difference in difference estimates and panel data:

Kenneth Scheve and David Stasavage (2012). “Democracy, War, and Wealth: Lessons from Two Centuries of Inheritance Taxation”. *American Political Science Review*.

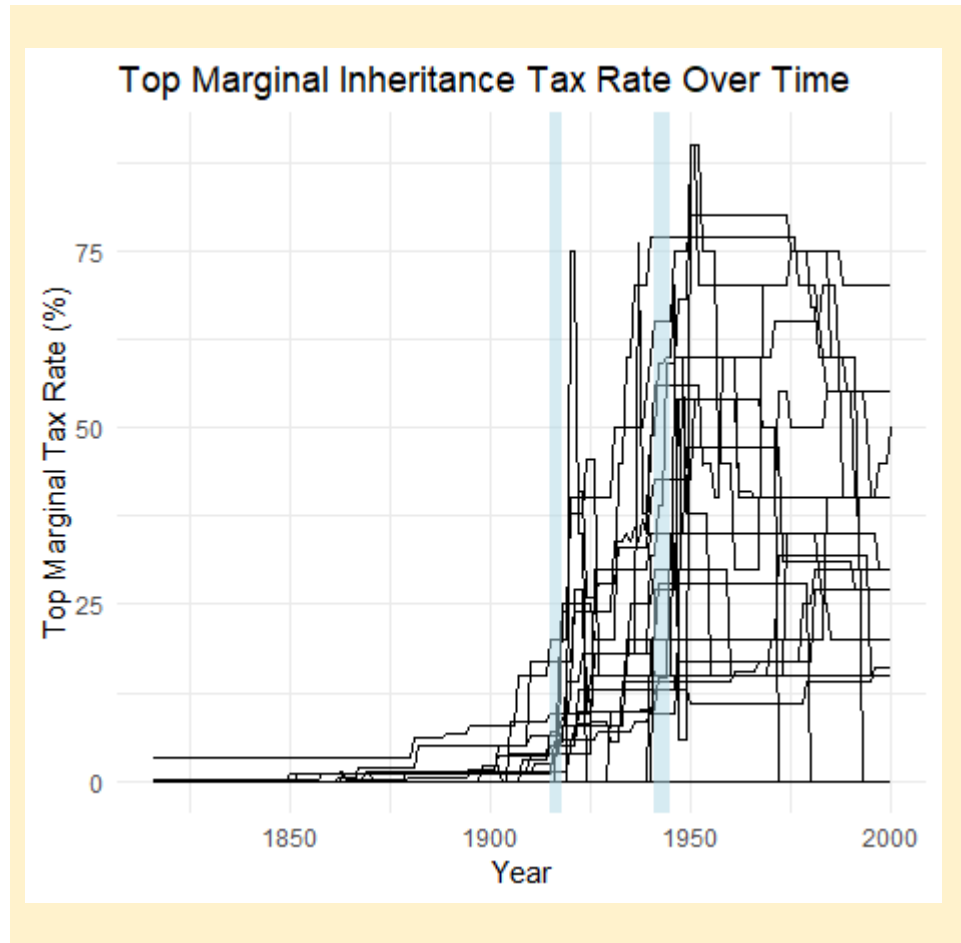
We will discuss this paper in class on Wednesday October 11th.

The key variables in this dataset, `ScheveStasavage_repdata.dta`, are:

Variable name	Description
year	Year of measurement
country, ccode	country, and associated code
topitaxrate2	Top marginal tax rate for a single direct descendant receiving cash inheritance, in percent units (so that a value of 50 indicates a 50% marginal tax rate)
himobpopyear2p	War Mobilization: engaged in interstate war, 2% of population serving in the military
unisuffrage	Universal Male Suffrage: all adult males eligible to vote in national elections

2.1

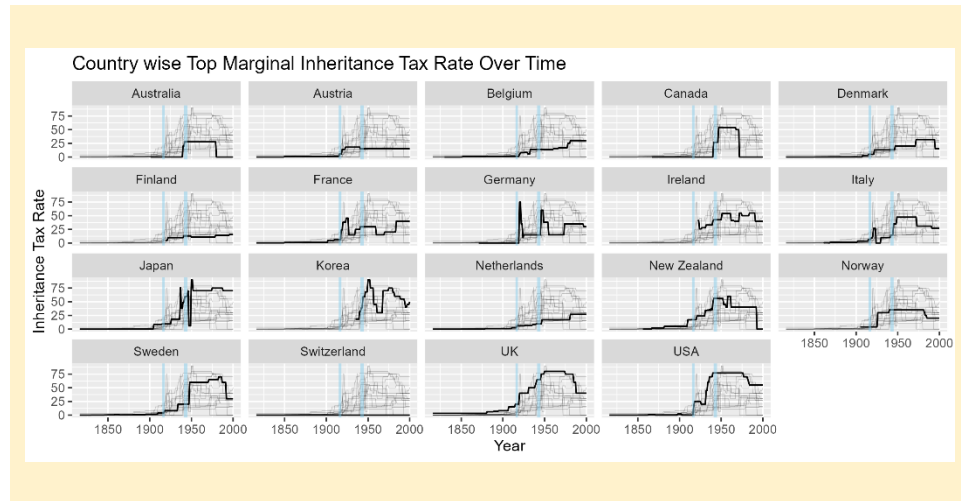
Make a line graph of a country’s top bracket inheritance tax rate for the whole dataset. Similar to Figure 2 in the Scheve and Stasavage paper, add a translucent shade for the periods of WWI and WWII. Countries entered the war at different times, but you may use the rough common periods, such as 1915-1918 and 1941-1945.



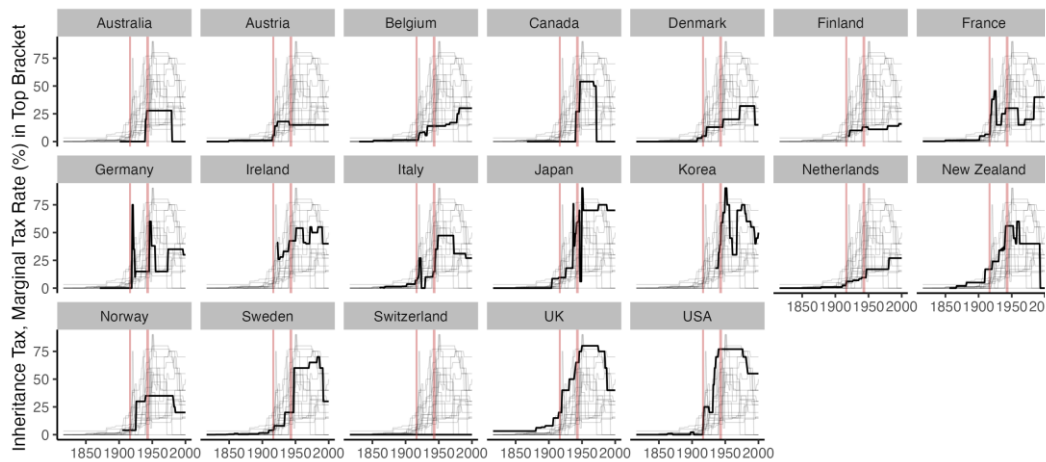
2.2

The graph you made too crowded to see apart each country. Scheve and Stasavage's Figure 1 makes separate small graph for three countries at a time, but this binning is arbitrary and limits the visual comparisons the reader could make. One effective way to make the graph clearer is to make a small multiples graph like the last graph [here](#), and shown below.

Make such a figure using `facet_wrap()`, where each facet represents a country, and save the figure using `ggsave()` as a png figure in a size that is not too small that the labels are squashed, but not too large that there is gaping whitespace. Paste in the figure here.



Optional but recommended: As shown in the example graph, under each country's trend line, show all the other country's lines in a translucent color so that the country's trend can be compared to all other countries.



Problem 3

3.1

Figure 2 in Scheve and Stasavage and the text around it discusses two countries as an example – UK and the Netherlands. Relate their discussion of these two countries in the context of their wartime mobilization hypothesis to the Card and Krueger minimum wage difference-and-difference design discussed in class. What is the treatment, who does it affect, and what is the outcome in Scheve and Stasavage?

The treatment in Scheve and Stasavage is the UK in post war period. Similarly in Card and Krueger, it represents those shops in New Jersey in the year post policy was implemented.

The outcome in Scheve and Stasavage which is the top marginal inheritance rate does increase during wartime. In the Card and Krueger minimum wage DID, the outcome which is the number of full-time employees in New Jersey after the policy did not reduce.

Due to the wartime mobilization of resources, the rich people in UK were taxed on inheritance and this affected them but by doing so it was argued to correct for some pre-existing unfairness in the way income had been earned. However, in Card and Krueger study, minimum wages did not affect employment numbers in stores located in New Jersey post the policy.

3.2

Report the simple average of the top inheritance tax rate in the 2 by 2 case explained in the previous question. That is, there should be four averages:

1. for each of the two countries (the UK and Netherlands),
2. before the two World Wars (defined here as 1914 and before) and after the wars ended everywhere (defined here as 1949 and onwards),

report their *average inheritance tax rate* in a well-labelled 2 by 2 table (in MS Word is fine here). Note you will dropping data in the interwar years 1915 and 1948 from the analysis.

Table 3.2 Average top inheritance tax rate		
Country	Pre-war	Post-war
Netherlands	0.4	21

UK	5	68
----	---	----

3.3

Recall that linear regression is a difference in means estimator, which means that its coefficients represent differences of subgroup means of the outcome variable.

To verify this, run the interactive linear regression representing the 2 by 2 difference in differences in problem 3.2, *represent the regression and the coefficients you are estimating in an equation, and* walk through how the *four coefficients reported can be constructed exactly from the means in 3.2.*

Note: Estimating equations are an integral part of a paper. When introducing an equation, you should follow the standard format along the lines of:

“I estimate the regression of the following form, <linebreak> _____
<linebreak>, where Y_{it} is ____, X is ____, and...”

I estimate the regression of the following form,

$$Y_{it} = \alpha_j + \gamma_t + \beta D_{it} + \varepsilon_{it}$$

Where

Y_{it} is the top marginal inheritance tax rate which is the outcome variable of interest

α_j is the country effect

The coefficient tells us that the difference between inheritance tax rate in UK and Netherlands prewar was 4.8 percentage points. In the table, it's the difference between the mean values of UK and Netherlands pre-war which is approximately (5- 0.4)

γ_t is the time effect

The coefficient gives us the time trend for Netherlands post and prewar periods which is 20.4 percentage points. In the table, it's the difference between average tax rate in Netherlands post and prewar: (21 – 0.4)

D is an interaction term that is 1 only if the country is UK and its t is post war period

β represent difference in means between D
= 1 and 0 conditional on country and time effects absent treatment

From the regression it can be inferred that the effect of slope on UK is 42.1 percentage points.

From the means in Table 3.2, it can be seen as the following: (UK post war – Netherlands post war) – (UK prewar – Netherlands prewar) = (68 -21) – (5- 0.4) = 42.1

And finally, the intercept coefficient from the regression gives us the baseline average for Netherlands pre-war which is 0.4 percentage. This is the value of Netherlands prewar in Table 3.2

Appendix

Problem 1

- The screencast at <https://vimeo.com/755382687> may be helpful, as are the help pages for `fixest` and `modelsummary`.
- This version will not exactly replicate the author's point estimates in Table 4 for the time being, because it lacks other control variables such as country fixed effects. We will return to that in a subsequent problem set.
- `coef_map` in the `modelsummary` output specifies the formatted name and order of the coefficients for the table. For example, for the regression $\text{lm}(Y \sim X1 + X2)$, `coef_map = c("X2" = "B", "X1" = "A")` would display the coefficients of X2 and X1, in that order, as B and A. Instrumental variables in `feols` adds a `fit_` to the fitted treatment in two stage least squares, but you will want to rename it the same as way as the OLS so it shows up on side by side in the same row.
- Recall the use of `gof_map = "nobs"` to show only number of observations.
- The paper has discussions of units and interpretation which will help you add confidence to your answer in 1.2.

2.1

- To draw separate lines for a variable such as `ccode`, map that variable to the `group` aesthetic. If you want to color the line by the variable, also map it to the `color` variable. For this problem, you do not need to color countries in different colors because there are too many countries for the coloring to be of any use.
- `ggplot` makes it easy to make visualizations with *variables*, but to make one-off annotations like a line or text we need a separate layer of *annotations*. Use the `annotate` function as a layer in your `ggplot`. See the examples at <https://ggplot2.tidyverse.org/reference/annotate.html>. In this problem, we want to create a rectangle (a `geom` of class "rect").
 - One tip that is not in the reference page is that you can use the values `Inf` to mean the max value in your graph, and `-Inf` to be the minimum value in your graph. So, `ymin = -Inf, ymax = Inf` sets the top and bottom of the rectangle to range the full scale of the plot.
 - The aesthetic `alpha` indicates the degree of transparency, where `alpha = 0` is completely transparent and `alpha = 1` is completely non-transparent. Because the value of transparency will be fixed across the data and will not vary by a variable, in this case we define it outside of the aesthetic mapping.

2.2

- Some general tips for visualizing time trends are at <https://clauswilke.com/dataviz/time-series.html>. For small multiples, see Wilke's chapter at <https://clauswilke.com/dataviz/multi-panel-figures.html>.

- For the optional feature to add a group of trendlines underneath, the trick here is to have *two* versions of `geom_line` geometries used in the same graph: One for the main line (one per each facet) and one for the group of other countries (some times called a spaghetti plot).
 - `facet_wrap(~ country)` will try to separate out all layers in the ggplot by the country variable, but if the facetting variable does not exist in the dataset, `facet_wrap` will replicate the entire dataset in each facet.
 - So, to make a background layer that shows all countries, you should provide a dataset that is modified so that it does *not* include the country variable.
 - Such a modified dataset can be provided by the `data =` argument in any geom (it assumes the main dataset by default). To drop a variable, use the minus sign within select (e.g. `select(scheve, -country)`).
- You can save your graph with the `ggsave` function (<https://ggplot2.tidyverse.org/reference/ggsave.html>). This allows for finer control for the width and height of your figure, and makes the graph replicable. A good size to start with is a 5 by 3 inches: `width = 5, height = 3`.

3.2 and 3.3

- We can use `case_when` again to make a variable that is `pre-war` (if year is earlier than 1914) or `post-war` (if year is later than 1948). The resulting variable can be put into a regression as a categorical binary variable too. In the regression, it would be useful to relevel the variable so that `pre-war` is the baseline category.
- Even though there are more than two years in the data, in this problem set we are working with the simplest 2 by 2 case. That means there should be only 2 types of time periods (`pre-war` and `post-war`) and two units (Netherlands and UK). In both problems 3.2 or 3.3, you do not need to consider the effect of different years beyond that 2 by 2 (e.g., the effect of 1913 vs. 1914). When we get to fixed effects regression, we will be accounting for these year by year differences.
- Well-labeled table in this class means: variables are labelled in formatted English so a reader can understand them (e.g., use “Post-War”, not “`post_war`” or “1”). Round digits in a way that a newspaper or published paper would.