ESM 245
Homework #2: Differences in differences

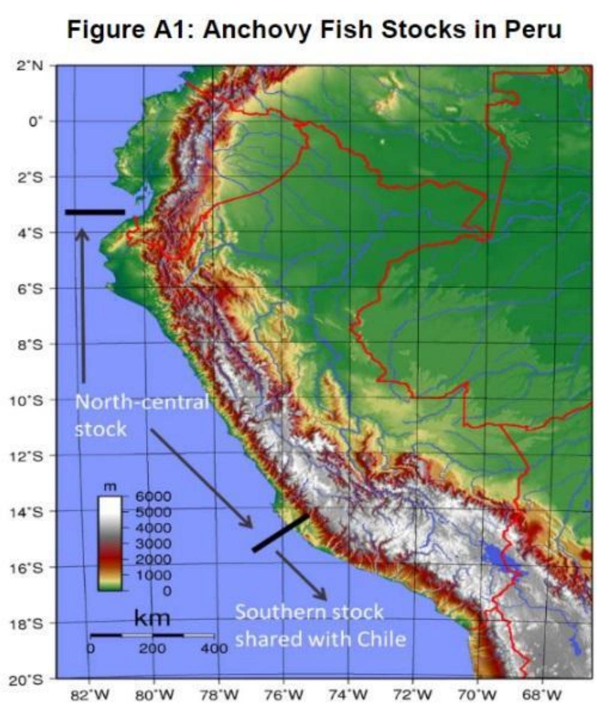
In this homework, you will conduct a differences-in-differences causal analysis of a fishery management reform in Peru. It is inspired by the setting in Natividad (2014), who studied the impact of implementing an IVQ (Individual Vessel Quota) for the Peruvian anchovy fishery. The objectives of this policy were to improve stock monitoring, reduce stock pressure in a short window of time (by lowering harvest per fishing trip through longer fishing seasons), and create incentives to avoid over-investment in competing for the resource. You will test whether the policy, which went into place in 2009, had a causal effect on fishery landings.

This homework will walk you through the actual calculations for differences-in-differences causal analysis. Please answer all the questions and keep your responses brief.

While the guided exercise employs R, you can follow the same steps in Excel, Stata, or other software with which you are proficient.

1. Setting and motivation

Our target species is the Peruvian anchovy, which has two fishing areas in Peru: the North Center zone and the South zone. In addition, there are two types of fishing fleets: 1. the steel fleet, managed by big corporations, and 2. the "Viking" fleet, managed by small fishermen's families. The species is important because it represents 90% of Peru's maritime landings, and 98% is exported as fishmeal (PRODUCE, 2019). In addition, this sector generates 12,000 direct jobs. Consequently, conserving the stock is crucial to keep production and employment in the coming decades.



Arias Schreiber and Halliday, 2013

Before 2009, the vessels harvested the resource until they met the total allowable catch (TAC). No individual caps existed.

In 2009, the Peruvian Government implemented a system of individual vessel quotas (IVQ) for the *North-Central* anchoveta fisheries[1]. This scheme also kept the TAC but distributed individual quotas to a limited number of vessels. Each vessel receives a percentage of the total allowable catch, which varies each season considering biological, weather, and climate variables. The historical catch of each vessel determines its respective TAC percentage. In addition, this scheme established a seasonal tax per harvested ton, satellite surveillance, and authorized docks to weigh the landings.

The Inter-American Development Bank (IDB) has hired you to evaluate the impact of the IVQ system on harvest. They suggest using the South area as a counterfactual area.

Data and packages

The Ministry of Production of Peru shared the dataset "df_peru.csv" with you. There, you can find the following variables:

- year
- month
- season: number of the fishing season (1 or 2)
- zone: fishing zone: S= South and CN = Center-North
- landing: Peruvian anchovy landing in metric tons

You will need the package tidyverse. You can use stargaze, kableExtra or other packages to show your tables.

2. Guided Diff-in-Diff

We are going to employ Differences in Differences to determine the **causal effect of the new policy on anchovy landings**. We will use the following functional form:

$$Landing_{tg} = \alpha + \beta_0 North_g + \beta_1 Post_t + \beta_3 North_g Post_t + \varepsilon_{tg}$$

where Landing_tg is the log of landings of the zone g during month t, α is the constant, North_g is a dummy for the North-Center region, Post_t is the dummy for the policy implementation, and North_g x Post_t is our difference-in-difference variable. Give some thought to why this difference-in-difference variable makes sense.

2.1. Read in the file **df_peru.csv**, and keep it in an object called df.
2.2. Create a new variable called **log_landing,** which is the log(landing) **.**
2.3. Create a dummy variable (call it **post**) to indicate when the treatment started (year 2009). In other words, the variable will be 1 starting in 2009 and thereafter.
2.4. Create a dummy variable (call it **north**) to identify the group exposed to the treatment (CN).
2.5. Create the **interaction** between the time the treatment started and the treatment group. Call this variable **did.**
2.6. Run linear regression. Use the functional form proposed above. Show the results in a table. You don't need to make a fancy table; it is enough if it is easy to read.
2.7. Interpret the did coefficient. Remember that we transformed landing by using log. This is your main result, so take a minute to really think about what this estimated coefficient is telling you.

---

[1] In the actual case, the policy was implemented one year later in the Southern zone.

3. Analysis

3.1. We used the South as a counterfactual region because the policy wasn't applied there. Describe one or more possible problems with this assumption.

3.2. To analyze whether the counterfactual makes sense, take the mean of log_landing before and after the policy for each zone (Remember Andrew's lecture on counterfactuals). Show your results in a table.

3.3. Does the assumption of parallel trends seem to hold? What are the implications of this for your analysis?