I hope you’re well. I finally sat down (having read my entire 204 textbook and am therefore more knowledgeable about econometrics) and devised a new model for the Ukraine situation.

My conclusions, in short:

* Control for GDP (general economic conditions) and ACLED battle deaths (conflict severity)
* We should use a **regression discontinuity model**, limiting the sample size to just countries that have one big airport that reopens after a while, and seeing how the OLS relationship changes after the intervention (airport reopening)
* We should use a **fixed effects design** too, if possible. Country fixed effects would allow us to model within-country changes only.

Explicit model:

Predicted FDI = β0 + β1*AirportReopensi*+ β2(year-C) + β3(year-C)\**AirportReopensi*+ β4lag(log(ACLED battle deaths)) + β5lag(log(GDP)) + country fixed effects

* Whereby C is the year the airport reopens

I have a comprehensive write-up attached detailing my thought process.

Let me know what you think, and I will get to work!

Best,

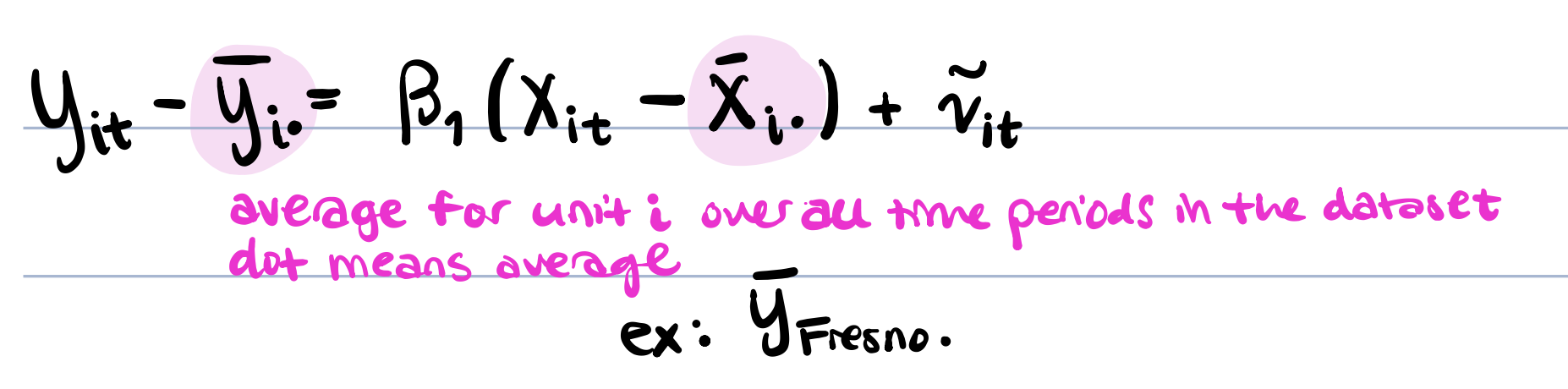
Michelle

* **Relationship of interest:**
  + How much does reopening an airport (controlling for conflict and economic conditions) increase foreign direct investment?
    - Relative to when the airport does not reopen
* **Regression discontinuity model:** 
  + Predicted Y = β0 + β1Ti + β2(X1i-C) + β3(X1i-C)\*Ti + β4X2 + β5X3
    - Explicit model:
    - Predicted FDI = β0 + β1*AirportReopensi* + β2(year-C) + β3(year-C)\**AirportReopensi* + β4lag(log(ACLED battle deaths)) + β5lag(log(GDP))
      * Whereby C is the year the airport reopens
  + Y = total FDI amount
    - (therefore we can generate how much the intervention increases total FDI in the year immediately after, and how the slope changes for the years following)
  + Ti = dummy for airport reopening
  + (X1i-C) = years before/after the airport reopens
    - C is cutoff is the year the airport reopens for this country
    - X1i is the year of the observation
    - (X1i-C)\*Ti is the interaction between the two: accounts for varying slopes before and after the cutoff
  + X2 = lag(log(ACLED battle deaths))
    - ACLED battle deaths used to approximate severity of conflict
    - Log used to standardize exponentially large numbers
    - Lag used because this would affect the FDI numbers the year *after*
  + X3 = lag(log(GDP))
    - GDP used to approximate multiple factors
      * Size of economy the year before
      * Not per capita: we want to talk about how much the intervention would add to the economy at large, not per capita
      * Makes the model more about how reopening the airport would affect Ukraine, given all economic conditions are the same
    - Log used to standardize exponentially large numbers
    - Lag used because this would affect the FDI numbers the year *after*
  + Probably also adding a fixed effect for countries (see highlighted below)
* **Endogeneity concerns**
  + Reopening the airport will be correlated with many things
    - Bring directly into the model:
      * Conflict severity (represented by ACLED battle deaths)
      * GDP level (represented by GDP from World Bank dataset)
    - Can’t bring directly into the model:
      * Underlying economic conditions, prior trade / economic activity before the war / right before the reopening
        + Including: trade relationships with other countries, foreign population, imports, exports, number of passengers flying into the country beforehand…
        + This can be approximated by airport size, which I have 2018 numbers for. However, this would introduce arbitrary variation into the model, because the intervention (airport reopening) could happen anytime in the period, and 2018 numbers would potentially absorb some of the post-treatment effect.
        + We would be able to use lagged passenger count, but this would similarly absorb true causal effect.

We may be curious about how reopening would increase passenger counts in Ukraine, but that’s a separate question we can study after this model.

* + - * + Therefore, we should be able to just approximate these variables with lag(GDP).

Unless we can think of a better variable.

* + Selection bias: planning on just looking at countries where we know that there is one primary airport and it reopens cleanly after some years
    - These countries probably had the conditions necessary, measured and unmeasured, to facilitate such a reopening
      * We can generally account for this by controlling for GDP and ACLED battle deaths
      * But this is not perfect. A synthetic control would have potentially been better, with more variables fed into the model, but for now this gets the job done.
* **Other variables we could have added but probably shouldn’t:**
  + See above for logic about airport size and # passengers flying into the country
  + Tax revenue as a % of GDP or as a total
    - I have data for this
    - Could show institutional quality, helping to model an important difference between opening in Ukraine versus other countries
    - But it may just show what style of economy plays into FDI increases
      * Could definitely add to model just to see. Maybe use the % since this would show more of the innate tax effect rather than controlling for GDP again.
  + Population
    - By controlling for GDP we approximate this already, also absorbed somewhat by ACLED battle deaths
    - We’re talking about adding to economy at large, not per capita
    - We can add this later if we want
  + Businesses in the country over time
    - I have data for this, ratio to population could show level of competition, entrepreneurial/grassroots engagement with the economy
    - I think this is not so important to introduce to the model, but we can add this if you think it is a good idea
* **Other models considered**
  + Synthetic control:
    - I couldn't get it to work last time, but even if I could, I think the RD is more parsimonious for what we want to test and would yield more interpretable results, potentially. I can try for this again later, though.
  + Difference-in-difference
    - There is no one time that the airport reopened for everyone. Therefore, the “control group” wouldn’t have a marker for when the treatment “would have been imposed but wasn’t” for them.
    - Limiting to just cases where there was a reopening wouldn’t solve things. We wouldn’t have a control group to calculate a difference to. This is why we have to use RD: it’s more flexible since you can just compare before and after a treatment for multiple discrete observations.
* **Concerns with this design**
  + For RD, we need to ensure there’s no weird clustering around the discontinuity
    - Because of the intervention happening in different years, maybe we should limit the window to just 3-4 years before and after (and widen the number of years captured with the data in general)
    - Test also whether all other vars are weird at the discontinuity
  + This would be estimating the local average treatment effect, or the immediate effect of the airport reopening, not really generalizable beyond this
  + Pooling data: in danger of producing endogeneity if conditions (contained in the error term) unique to a specific country would affect FDI
    - **It would be a good idea to include fixed effects for countries** 
      * Would be N=9ish for each country, depending on how many years I include.
      * The controls of GDP and ACLED would then soak up any endogeneity innate to just within-country variation, which is still very relevant!
    - **Adding a fixed effect to the model:**
      * De-meaned approach?: subtract unit-specific average values from independent and dependent variables
        + 
        + Could also do the dummy approach.
        + I will figure out how to code this in R.

Also how to incorporate for all variables.

* + - * Dangers:
        + Could inflate standard errors if irrelevant, but we think that these fixed effects would be non-zero. Therefore we should control for them.
        + We would not be able to estimate effects for stationary variables, but we don’t have any of these in the analysis.
      * We could also do two-way fixed effects, but this may soak up some of the changes in the post-treatment variable.
        + time fixed effects are less important but would allow us to control for year-specific variation
        + Time fixed effects may be too correlated with the post-treatment variable, i’m not sure….