REPLICATION EXERCISE 2: MIGUEL AND ROLAND (2011)

VOJTĚCH BARTOŠ

Joint classroom discussion

In this exercise we replicate the results of the paper by Miguel and Roland (2011) (MR): The long-run impact of bombing Vietnam. We will only work with the district level data for the sake of time.

(1) Open the MR paper and read the theoretical framework in Section 2.2 discussing the neoclassical growth model with a poverty trap component generated by a subsistence consumption component added to the standard Solow model we discussed in class. (30 min)
   (a) Comment on the difference between the standard model we discussed and the extended model presented in MR.
   (b) Comment on how do minimum consumption level c_{min}, population growth n, and depreciation rate δ affect the poverty trap level k_{trap}.
   (c) Why is Vietnam a good case to study the existence of poverty traps?
   (d) Why do the authors focus on both the district level and the province level? Why not only focus on the smaller units, the districts?
   (e) How do MR support the argument that capital loss due to bombing is much larger than the loss of population. Why is this important for the workings of the theoretical model.
   (f) In the model presented, what would happen to districts affected by bombing if cross-district factor mobility is not allowed?
   (g) If full factor mobility is allowed, would private individuals invest in a region where $k_{it} < k_{trap}$? Would a central planner (a government) want to invest in this region? Explain the logic in words or using the model parameters.
   (h) If full factor mobility is allowed and if only $m < n$ regions are affected by bombing (where n is the total number of districts in the country), under what conditions would the country experience a poverty trap?

Individual team work

(1) To get a better idea about the historical content of the Vietnam War (although we talked about it somewhat in the lecture), read Section I, Introduction. (10 min)
(2) Download the dataset *MR2011_replication.dta* from my website. Open the dataset in Stata and explore the data. (10 min)

Joint classroom discussion

We'll wonder how the intensity of bombing during the war affects current day economic development in Vietnam on the district level. Among other variables, the database also contains previously classified data on all ordnance by type used in Vietnam by US and allied Air Force and Navy between 1965 and 1975. The authors aggregated this data on a district level.[1]

(1) How can we use the data to assess the damage done to the respective districts? What would be the ideal data that you would want to have in order to answer the question of how does a shock to capital intensity (war, in our case) affect economic growth? (5 min)
(2) Under what assumption would the missing access to data on ordnance used by the Army ground forces (note: we only have data for Navy and for Air Force) lead to the presumed negative effect of bombing on economic performance downward? (5 min)

Individual team work

(1) Replicate Panel A of Table 1 describing the summary statistics of US ordnance used.[2] (25 min)
   (a) You will need to create two new variables:
      (i) ordnance: "total US bombs, missiles, and rockets" (use the following variables: General_Purpose Cluster_Bomb Missile Rocket Cannon_Artillery).
      (ii) ordnancekm2: "total US bombs, missiles, and rockets per km^2"
   (b) Which district (districtname) in which province (provincename) received the lowest amount of ordnance per km^2? Which district the highest amount? How much ordnance was used in the following areas:
      (i) province of "Ha Noi (City)"
      (ii) province of "Ho Chi Minh (City)"
      (iii) province of "Quang Tri".
      (iv) Further, how does the amount of ordnance per km^2 differ in areas above and below the 17^{th} parallel (North latitude), the border between North and South Vietnam?
   (c) Now plot the graph with North latitude on the horizontal and the ordnance per km^2 variable on the vertical axes. What do you observe?[3]

[1] These are the variables of interest (using data editor, look at the descriptions to understand what these variables stand for; not all are explained, though): Ammunition Cannon_Artillery Chemical Cluster_Bomb Flare Fuel_Air_Explosive General_Purpose Grenade Incendiary Mine Missile Other Rocket Submunition Torpedo Unknown UnlabeledUSAF A AAC AC ACC AP COM COMM CVT HC HCC HCP HCPD HCVT HE HECVT HEPD HP HVTF ILL ILLUM ILUM MK MK07 MK10 MK12 MK7 MK70 MK8 P P0 RAGON RAP SHRKE UNLABELEDUSN VC VT VTN VTNSD VTSD W WP

[2] Hint: for correlations use the command `pwcorr` (the simple correlates will differ from the amounts; you should weight by the total population in the district to get the numbers used in the paper)

[3] You can have a look at the map of Vietnam for reference.

(d) You can see that there are many variables on ordnance that remain unreported in the paper. Let's generate a new variable that combines all the variables on ordnance listed above and call this variable allordnance by simply summing them up. Generate another variable allordnancekm2 that reports all ordnance used per km^2.
   (i) How much ordnance per km^2 remains unreported? Could this be a problem? Explain.
   (ii) Now examine the correlation between allordnancekm2 and ordnancekm2 (again, weight by district population). Would you change the answer to the previous point? Why?

(2) We have seen that there is quite some variance in the extent of bombardment in respective districts. Given the predictions in the model, if the bombing is severe enough, we might observe a permanent effect (a poverty trap) on economic outcomes in the future. Let us explore this in a regression analysis. As a main variable of interest that we will use as a proxy for the destruction of physical capital will be the variable ordnancekm2.

Run a series of OLS regression: $y_i = \beta_0 + \beta_1 ordnancekm2 + \varepsilon_i$ (15 min)

(a) y_i stands for the following variables:
   (i) poverty rate: `poverty_p0`
   (ii) population density in 1999: `popdensity1999`
   (iii) physical capital investment levels: proportion of households with access to electricity in 1999: `elec_rate`
   (iv) human capital investment levels: proportion of literate respondents in 1999: `lit_rate`
   (v) (they also study consumption data but only on the province level, we'll skip this)

(b) Store the results using `outreg2` function

(c) What results would you expect based on the model prediction should the poverty trap story be true? Comment on the results you obtained. How do you read the coefficients?

(d) MR also comment on differences between the effect of bombing on poverty in the former North and South Vietnam (variable south), and between mostly rural and mostly urban districts (as defined by popdensity6061 above and below 200), they also examine non-linear effects of bombing by including a squared term of ordnancekm2 in the regression, and instead of using the continuous variable ordnancekm2, they rather generate a dummy variable for top 10% of districts that were most affected (use command `xtile` to create percentiles and generate a dummy for the tenth percentile); Run all these regressions and comment on the differences from the aggregate regressions.

(e) What is the problem with running this regression in terms of identification?

(3) Now you see that the MR control for many variables in their regression. Enrich your model for the following variables to reproduce the column 2 in Table 4: (10 min)

   (a) Population density (province) in 1960-61 (divided by 100): popdensity6061
   (b) Former South Vietnam dummy: south
   (c) Proportion of land area 250500m: area_251_500m
   (d) Proportion of land area 5001000m: area_501_1000m
   (e) Proportion of land area over 1000m: area_over_1000m
   (f) Average precipitation (cm): pre_avg
   (g) Average temperature (Celsius): tmp_avg
   (h) Latitude (N): north_lat
   (i) Soil type (18 categories): soil_1, soil_3, soil_6, soil_7, soil_8, soil_9, soil_10, soil_11, soil_12, soil_14, soil_24, soil_26, soil_33, soil_34, soil_35, soil_39, soil_40, soil_41
(4) Comment on the reason for including the control variables in the model. You can find some supporting arguments in section 3.3. (5 min)
(5) If you did everything right, the point estimates of your regressions should exactly match those in column 2 in Table 4, but the standard errors you estimated are smaller in your case than in the case of the MR paper. The reason is that in the regression you do not account for correlation of disturbances within specific clusters. Run the same regression again and cluster the standard errors at the province level. Just in case, also use robust standard errors to correct for heteroskedasticity. (15 min)
   (a) Do your results now match those in the paper? Save the results using `outreg2` function.
   (b) If the coefficients match the data in regression tables, reproduce also the results in columns 3 and 4 (hint: look at the variable provincename) in Table 4 and save the results using `outreg2`.
      (i) What is the reason for dropping Quang Tri province? Examine the extent of bombardment in this province relative to that in the rest of the country. Comment.
      (ii) Note: Why are the variables south and popdensity6061 missing in column 3 Table 4?
(6) Although weak, we see that there is a negative relationship between historical bombing and current poverty rate. Why can't we make a causal link between bombing and poverty rates and merely talk about correlation? (5 min)

(7) MR find an instrumental variable that resolves the issue you just described. It is the absolute distance from 17^{th} Northern latitude, the administrative border between North and South Vietnam. (20 min)
   (a) What needs to be satisfied so that this is a valid instrument (comment on assumptions on 1) the border distance and bombing and 2) the border distance and current poverty rate)?
   (b) Let's now do the IV regression where we instrument for the intensity of bombing by the absolute distance from the 17^{th} Northern latitude.

(i) First, generate the variable abs_dist_17 that contains a distance from the 17^{th} Northern latitude (absolute value).
(ii) Now we should check if our intuition was correct and whether there is a link between bombing intensity and the absolute distance. Regress ordnancekm2 on abs_dist_17 and the rest of the controls you used to replicate column 2 in Table 4. Cluster at the province level and use Huber-White (robust) standard errors. This should give you the same results as column 2 in Table 3. Does it? Comment briefly.
(iii) Use the `ivreg` command in Stata (see help for syntax). Use abs_dist_17 as an instrument for ordnancekm2. Otherwise, use the same specification you used to replicate column 2 in Table 4 and treat all other variables as exogenous. Again, cluster on the province level and use the Huber-White (robust) standard errors. Did you get the same results as MR? Comment briefly.

(8) If there is still time remaining, replicate the rest of the results on district level: columns 2 to 6 in Table 7, columns 2 to 6 in Table 8, panel A of Table 9.