Problem Set 2

This problem set is due February 27th. Upload your write up and any code files to the Github Classroom before midnight. You may work together, but you must turn in separately written (unique) write ups and/or code.

In this problem set, we are going to consider the relationship between economic growth and civil war. We will base on study on data and analysis from Miguel, Satynanath, and Sergenti (2004). A copy of this paper is included along with a code book describing their data. The dataset is found in the file `conflict.dta`.

1. Let's start on the theoretical end of this relationship and consider the following model for the relationship between civil conflict w_{it} and economic growth g_{it}

$$w_{it} = \alpha_i + \tau_t + \beta_1 g_{it} + \beta_2' x_{it} + \varepsilon_{it} \tag{1}$$
$$g_{it} = \kappa_i + \rho_t + \gamma_1 w_{it} + \gamma_2' x_{it} + u_{it}, \tag{2}$$

   where x_{it} is some length-k vector of control variables. Additionally, let's add some structure in the sense of imposing weak exogeneity

$$\mathrm{E}[\varepsilon_{it} x_{it}] = \mathrm{E}[u_{it} x_{it}] = 0$$

   and assuming that each unit i is independent of the others. For the within-estimator to be a consistent estimator of β_1 in Eq. 1, we will need to establish that $\mathrm{E}[\varepsilon_{it} g_{it}] = 0$.

   (a) Rewrite equations 1 & 2 to be in reduced form. This means that w_{it} and g_{it} should be functions of only the exogenous variables x_{it}, composite parameters made up of combinations of the above parameters, and a composite error term. Another way to say this is to solve the pair of equations for w_{it} and g_{it} such that they only appear on the left-hand side.

   (b) Find the expected value of $\mathrm{E}[\varepsilon_{it} g_{it}]$. What additional assumptions would be required to make it 0? Are any of them believable or plausible? Can the within estimator provide a consistent estimate β_1 if fit to Eq. 1?

(c) Suppose you fit the reduced form equations where you regress w_{it} on just x_{it} and g_{it} on just x_{it}. Could you use these regression coefficients to solve for $\hat{\beta}_1$? Why or why not?

(d) Let's separate x_{it} and it's original parameters as follows

$$\begin{aligned} x_{it}' &= (z_{it}, q_{it}')' \\ \beta_2' &= (\phi_0, \phi_1')' \\ \gamma_2' &= (\psi_0, \psi_1')' \\ \beta_2' x_{it} &= z_{it}\phi_0 + \phi_1' q_{it} \\ \gamma_2' x_{it} &= z_{it}\psi_0 + \psi_1' q_{it}, \end{aligned}$$

where z_{it}, ϕ_0 and ψ_0 are scalars and q_{it}, ϕ_1 and ψ_1 are length-$k-1$ vectors. Rewrite your reduced form equations with these substitutions.

(e) Let $\phi_0 = 0$. Can you consistently estimate β_1 using a function of the reduced form estimates now? Why or why not? (**HINT:** The answer is yes, but you have to explain how). What does this tell you about what empirical moments identify $\hat{\beta}_1$? (i.e., what characteristics of the data actually determine $\hat{\beta}_1$?

(f) Note that when $\phi_0 = 0$, z_{it} is an instrument for g_{it}, explain why it is valid and relevant.

(g) Based on all of your insight above, we decide to fit the reduced-form system equations:

$$y = \mathbf{X}\begin{bmatrix}\lambda \\ \delta\end{bmatrix} + D_N\begin{bmatrix}c \\ k\end{bmatrix} + D_T\begin{bmatrix}s \\ r\end{bmatrix} + e,$$

where

$$y = \begin{bmatrix}w \\ g\end{bmatrix} \qquad D_N = \begin{bmatrix}I_N \otimes 1_T & 0 \\ 0 & I_N \otimes 1_T\end{bmatrix}$$
$$\mathbf{X} = \begin{bmatrix}X & 0 \\ 0 & X\end{bmatrix} \qquad D_T = \begin{bmatrix}I_T \otimes 1_N & 0 \\ 0 & I_T \otimes 1_N\end{bmatrix}$$

Here, $c = (c_i)_1^N$, $k = (k_i)_1^N$, $r = (r_t)_1^T$, $s = (s_t)_1^T$, λ, and δ are the reduced-form parameters that are composed of the parts you described above. Likewise, e is combined reduced-form error term,

which you also described above. Let $\hat{V}$ be an appropriate estimated covariance matrix of $(\hat{\lambda}, \hat{\delta})$, find analytic expressions for $\hat{\beta}_1$ and its standard error from the reduced-form estimates found by fitting the system model. (**HINT:** Use the delta method to find the the standard error)

2. Let's turn to the data. For this problem you may use any packages you and pre-canned routines you want except where noted. I believe the main variables they use are as follows:

   - `any_prio`: Is there an active civil conflict in country i, year t (0/1)? This comes from the PRIO database. An active conflict is one with at least 25 battle deaths per year
   - `gdp_g`: Annual economic growth rate in country i, year t
   - `gdp_g_l`: Annual economic growth rate in country i, year $t-1$
   - `y_0`: Logged GDP per capita in country i in 1979
   - `polity2l`: Polity 2 score in country i, year $t-1$
   - `ethfrac`: Ethnolingustic fractionalization. Probability that two random citizens from country i are from the different ethnic groups. Based on Soviet ethnographic data.
   - `relfrac`: Religious fractionalization. Probability that two random citizens from country i are from the different religious groups. Based on CIA factbook.
   - `Oil`: Is more than a 1/3 of country i's export revenue from oil?
   - `lmtnest`: The logged proportion of the country that is covered in mountains
   - `lpopl1`: Logged population of country i at time $t-1$

   (a) Using the above list, verify that that we've identified the right measures by replicating the following parts of Table 1
      - The first row of section A
      - Section C
      - All but the last row of section D

   (b) The authors are interested in the effect of economic growth on conflict. Do you think a specification where we regress `any_prio` on the other variables listed makes sense? Why or why not?

(c) Reproduce columns 2-4 of Table 4 as best you can. For each of these interpret the effect of growth at time t on conflict. Which model specification do you think is the most convincing? Why or why not?

(d) Refit the model from column 4 of Table 4 with two-way fixed effects instead of the country-specific time. Compare the standard errors from clustering on country, clustering on time, Driscoll-Kraay, and two-way clustering. What differences do you observe? Which you use if this was your paper. Justify your answer.

(e) The authors propose using rainfall as an instrument for growth with the variables `GPCP_g` and `GPCP_g_l` measuring the average growth in rainfall in country i at times t and $t-1$, respectively. Briefly describe what we need to assume for these instruments to be valid and relevant for the two growth measures. Do you think these instruments are valid? Why or why not?

(f) Use 2SLS to fit a model where we regress `any_prio` on `gdp_g` using `GPCP_g` as an instrument. Control for democracy and population and use two-way fixed effects. Verify your analysis from question 1 by then using a system estimator and delta method to get to the same estimate for the effect of growth on conflict.

NOTE: We want to make sure that we have different fixed effects for the equations, *but* we don't want the standard errors to be over confident by double counting the number of units. An easy way to avoid problems is to create new country and year variables that are equal to `ccode` and `year` for the part of the data that is used in equation 1, and equal to that plus 500 (or some other large enough number) otherwise. Then cluster on the original `ccode` variable (or `year` or both). If you've done this right the estimates and standard errors should match the results of fitting the equations separately (subject to small differences in degrees of free in the standard errors).

(g) Having satisfied yourself that we understand the identification of this approach, go ahead and attempt to reproduce columns 5-6 from Table 4. Interpret the effect of growth on conflict.

(h) Consider the first stage regressions of rainfall on growth. Are these strong instruments?

(i) Consider an overidentified model where `GPCP_g` and `GPCP_g_l` instrument for just `GPCP_g` (i.e., drop `GPCP_g_l`). Fit this model with time-varying controls and two-way fixed effects. Fit the model with GMM using a weighting matrix and covariance matrix that you think is appropriate, justify your answer. Interpret the results and test the hypothesis that instruments are valid.

(j) Finally, suppose we were really interested in the effects of democracy and ethnic fractionalization on conflict. Choose a set of controls and fit your model with the following estimators

- A fixed effects estimator that allows for time-invariant effects (e.g., a Hausman-Taylor-like estimator, with or without random effects)
- Munlak (with or without the random effects)

Interpret your results. What if anything do you find?