How Democracy Affects the Growth of Nations?

Enrique Moral-Benito*

January 2009

One of the most important questions in economics is, what causes economic growth and thus prosperity for the people of the world? From an empirical point of view, since the seminal work of Barro (1991), a huge number of papers have claimed to have found one or more variables correlated with the growth rate. More concretely, in this lesson we will try to get some conclusions about the effect of the level of democracy on the growth rate across countries.

Empirical framework

The empirical analysis of growth across countries is mainly based on the extended version of the neoclassical growth model that you have probably seen in Macro I with Josep. In equation form, the model can be represented as:

$$\gamma = f(y, y^*)$$

where γ is the growth rate of per capita GDP, y is the level of per capita GDP, and y^* is the long-run or steady-state level of per capita GDP.

The target value y^* depends on an array of variables but there is no consensus about which are these variables. This raises an important question: which regressors should I include as control variables in growth regressions? This causes that such control variables depends on the particular study and the author's criterion.

In 1997 Xavier Sala-i-Martin ran two million OLS regressions. His main objective was to account for all the possible combinations of control variables when analyzing the effect of 59 different variables on growth. Our focus here is less ambitious since we only want to have a look to the effect of one variable, the level of democracy, on growth. For this purpose we will use the dataset that Sala-i-Martin compiled for his exercise.

The data

The data is available at http://www.columbia.edu/~xs23/data/millions.htm. You can download an excel file and then, simply copy and paste it into the Stata's Data Editor. The dataset is a

*E-mail: enrique.moral@gmail.com

cross section of 134 countries that comprises 62 potential growth determinants plus our dependent variable, the growth rate γ. In the excel file you will also find a list of the variables.

The level of democracy in a given country is proxied by the variable political rights (labeled as **x32** in the data file), which is an index that ranges from 1 to 7, where 1 corresponds to the highest level of democracy and 7 to the lowest.

Problem 1 *Since the ranking is a bit counter intuitive, the first thing we could do is to construct in Stata (using the* `"gen"` *command) a new index from 0 to 1 where 0 is the lowest level of democracy and 1 the highest.*

Problem 2 *Then, we can have a look to some variables we have, how many observations for each one, whether there are outliers and missing values or not... we can either open the data browser or get some descriptive statistics for this purpose (Hint: you can use the command summarize). On the other hand you can also create some graphs using the drop down menus, for example a scatter plot of growth against democracy. What does it look like?*

Running some regressions

Once we have become familiar with our data, we can further investigate by running some regressions. Depending on your curiosity, you can run billions of regressions looking for different effects.

Problem 3 *Given the title of the lesson, the first thing we could do is to run a regression of growth on democracy. What do you obtain? Can we conclude that democracy positively affects growth?*

Problem 4 *Could you replicate the obtained estimates and standard errors using Gauss instead of Stata?*

Omitted variables bias (OVB)

Note that in *Problem 3* we have just run the following regression:

$$\gamma_i = \alpha + \beta \; democ_i + u_i$$

where $democ_i$ is the level of democracy in country i. In this regression, u_i can be interpreted as all the "unobservable" factors that affect growth (γ_i) besides the democracy level. If there are some observable factors not included in the regression, then they are implicitly included in the term u_i. Moreover, as you have seen with Stephane, if these factors are correlated with the democracy level (i.e. $cov(democ_i, u_i) \neq 0$), the OLS estimate of the effect of democracy on growth (β) will be inconsistent and biased, this is known as the omitted variables bias.

Getting rid of OVB

The idea is to include the observable omitted variables in the regression and thereby estimate the effect of one regressor (democracy) while holding constant the other variables.

For example, one could easily argue that the level of GDP will affect the growth of GDP (since we did not explicitily include it in the first regression, we can interpret that it was in the u_i term). Moreover, richer countries are typically those with higher levels of democracy. If both arguments hold, then, the level of per capita GDP is one important factor omitted from the first regression that will lead to a bias in the OLS estimate. Therefore, per capita GDP should be included as a control variable when estimating the effect of democracy on growth. This is also illustrated by the empirical framework presented above and derived from the neoclassical theory of growth.

Problem 5 *Repeat the regression in Problem 3 but including per capita GDP (labeled as **x1**) as a control variable. Does the estimated effect change? Can you argue the existence of more relevant omitted factors in this context?*

For example, we can argue that the more educated the population in a country, the higher the growth rate of that country. Furthermore, higher democracy leads to a better public education system that causes the level of human capital in the form of education to increase. Given the above, we should control for education in our regression in order to get an unbiased and consistent estimate of the effect of democracy on growth when holding constant education and per capita GDP.

Problem 6 *Repeat the regression in Problem 5 but including also the primary-school enrollment rate (**x2**) as a control variable. What happens with the estimated effect now? What should we conclude?*

Problem 7 *The variable **x55** in the dataset contains the percentage of jews in the total population of the country. What happen if you include as a control variable instead of education? why?*

Going further

When we are involved in an empirical research, it is extremely important to think carefully about the particular circumstances of the issue at hand. For example, in our current investigation we may think that some amount of democracy is better than none, but once a certain level of democracy has been reached, an additional increase in the democracy level has no (or even a negative) effect on growth.

Problem 8 *How would you test this hypothesis?*

Some final caveats

- In the growth literature, as the number of countries in the world is limited, the number of available observations is also limited. In our case, we had more or less a hundred of countries. This of course is an important limitation of this literature, since the degrees of freedom is typically small. Therefore, the results are usually sensitive to changes in the empirical specification. This problem is not present for example in the microeconometrics literature where you may have a panel of 10,000 households.

- Another worry is the lack of anonymity. When you work with countries, you more or less know what these countries are and what would you expect in some contexts. For instance, you know that by dropping all OECD countries from your sample, your results may change a lot. This could lead us to the problem of data mining.

- With the inclusion of more control variables in the regression, as we have just done, we can alleviate the omitted variables bias to some extent. However, as long as we have some unobservable country-specific characteristics such as the ability of its population (or the country's historical background) correlated with one or more of the right hand side variables, this bias will be always present. This is so because these unobservable characteristics affect not only the growth rate but also other growth covariates such as the level of education. Since these characteristics are typically very hard to measure, how can we control for them and eliminate the omitted variable bias they generate? The most promising solution is the consideration of country-specific fixed effects in the error term by resorting to panel data techniques.

- Finally, working with this kind of cross sectional data in the growth context may very probably derive in endogeneity biases of the OLS estimator. This is so because in the kind of regressions we have run, we are not sure whether it makes more sense to regress growth on democracy or democracy on growth. That is to say, if we find a positive partial correlation between two variables (e.g. education and growth), does this mean that higher education cause higher growth? or vice versa? This "simultaneity" problem will of course make OLS estimates misleading and biased. One solution to this problem could be the use of instrumental variables (you will see more on this with Stephane).