

THE EFFECT OF INCOME INEQUALITY ON GDP GROWTH IN TURKEY

I.Motivation

Income inequality among citizens within a particular country is becoming more and more problematic as countries focused on maximizing their GDP growth, whatever the result is. They don't care about income inequality or climate change or even their identified overheated economical results (as we're suffering in Turkey). In the name of GDP growth, countries can cancel out agreements on productivity or negative externalities which can cause climate to change, countries can put barriers and make all the world's economy suffer from this. In such economical and political atmosphere of the world, we decided to see the effect of Gini coefficient on GDP growth, and we wanted to see if it's really worth suffering from high Gini in order to have high GDP growth. We think that in terms political, economical goals of countries, Gini should be one of the concerns they must care about. Because overall equal welfare and happy people are also important for our future, talking simply.

In this research, our point of interest was mainly the effect of Gini coefficient on GDP growth. We tried to add control variables to hold constant potential candidates for the omitted variable bias. In addition to Gini(gini), we added Domestic Gross Savings(gross), Population Growth Rate(poprowth), Unemployment Rate(unemprate) in our regression to see the real effect of Gini on our dependent variable, on average.

II.Literature Review

There are some papers indicating that increase in Gini results in lower GDP growth. These results are rather parallel to our result. For example, in one text in OECD's website, it's clearly indicated that %1 increase in gini causes GDP growth to decrease by %0.6 and is statistically significant. There are also other papers stating that higher income inequality doesn't hurt GDP growth. In addition to them, there exists such papers that don't reach a clear and significant result. For example, one paper indicates that country's wealth plays a big role whether high Gini is good for growth or not. If we dig more, we know we can reach many different views and many different focuses on this topic.

III. Data

In spite of all the researches we have done so far, there are problems to dig such data for Turkey. The data is more available for the recent years, whereas it's rather difficult to find the data relating to earlier years. Data was either missing or not reliable. But we think that we found sufficient reliable data to run our regression.

We used many data sources like FRED(fred.stlouisfed.org), The World Bank, some research&report papers, , .

We wanted to control for such variables that may cause omitted variable bias. Our purpose was holding constant all potential variables that may have effect on our dependent and independent variable and see the real effect of a change in Gini on GDP growth on average. We were careful not to add variables that may cause reverse causal effect. Gini, gross, unemprate, popgrowth are not included the equation of GDP(C+I+G+NX) so that there is not reverse causal effect. It is important to add unemprate, popgrowth or gross in our regression in addition to gini because they all seem to have correlation with gini. And these variables are rather important measurements that affect economical activities.

Expected signs of coefficients are: (-) for gini, (+) for popgrowth, (+) for gross, (-) for unemp

Here is the all statistical data before we start.

Variable	Obs	Mean	Std. Dev.	Min	Max
gdpgrowth	56	4.855045	3.949045	-5.962311	11.21282
popgrowth	56	1.881892	.4128249	1.203624	2.407683
gross	44	25.29813	5.252922	18.19062	38.40107
gini	23	.4430957	.052819	.38	.56
unemprate	56	8.096429	2.202227	3.4	13

After dropping all missing values,

Variable	Obs	Mean	Std. Dev.	Min	Max
gdpgrowth	20	5.349208	4.325649	-4.704466	11.1135
popgrowth	20	1.58747	.3005477	1.203624	2.251607
gross	20	24.23807	4.362168	18.19062	38.40107
gini	20	.42856	.0380993	.38	.52
unemprate	20	10.055	1.229238	7.9	13

*We realized that there is no difference if we drop missing values or not, when we regressed them Stata did all adjusting by itself**

IV. Methodology

We also thought to run regressions in log terms but these variables are already given by percentage terms, so It's not very good idea to run them in log, so by econometrical method we used time-series data and did all the regressions in computer software that is Stata which we can implement our computational methods in. We used level-level regression. Our limitation in this empirical methodology is that we can't observe the change in independent variables regarding to their old value, rather we are taking into account the percentage change in our overall value of interest. For example, %1 increase in gross means that Gross Saving is %1 more of the GDP. But from %40 to %41 change means %0,025 change comparing to it's previous value(%40). Actually, it's not a limitation, we need the %1 change in our value

of interest. In addition to that biggest limitations that there are 33 missing values of Gini and 3 missing values of gross. It is very difficult to find Gini values for Turkey. Also, it was not very easy to rely on the data we had found, we had to enlarge our extent of search and spent hours to be less suspicious.

As we indicated above, we assumed our data to be homoskedastic(using robust term) and to assume our betas to be unbiased and consistent, we assumed that $E[u/x]=0$, there is no selection bias, our dependent and independent. randomly&independent&identically distributed, large outliers are unlikely. We need large sample to have our coefficients variables as small as possible. But our data is not really large, especially caused by Gini's missing values.

V. Results

(1) First of all, we ran the regression gini on gdpgrowth(robust), what we found was:

Coefficient of gini: -13.79997	p-value: 0.421	Prob > F = 0.4205
Coefficient of constant: 11.59692	p-value: 0.137	R-squared = 0.0311

In this regression, we found that p-value is not small enough for gini to be statistically significant so we can't regret that the coefficient of it to be zero. In addition to that prob>F is also larger than 0.05, so it supports our p-value's result. Also R-squared is so low that it is not logical to run this only with gini.

In addition to that, when we added „robust” term to our formula to make sure that it satisfies the homoskedasticity, we saw that results not even changed. From now on, we've applied the robust formula in all our regressions.

(2) So, we wanted to add control variable. We added unemprate to see the effect. Unemprate seemed logical to add because there seems correlation between unemprate and gini.

Coefficient of gini: -47.4869	std: 21.58586	p-value: 0.040	Prob > F = 0.0997
Coefficient of unemp: -1.036548	std: 0.5000139	p-value: 0.051	R-squared = 0.1252
Coefficient of constat: 36.28051	std: 13.52131	p-value: 0.014	
Corr(gini, unemp): -0.5595	corr(gini, gdpgrowth): -0.4062		

After adding unemp, we see that unemp almost turned out to be significant, and gini also seemed significant. We saw that there was positive bias in the first regression that we ran only with gini, but now it decreased. The bias was positive. We can conclude this by the correlation of gini and unemprate(also unemprate's coefficient is negative). Their product is positive, so it's compatible with our result(*later in our regressions we thought not to add unemp in our regressions so gini's actual sign became much more smaller than this, making everything more complex and deeper research actually is needed into it(4)*)

Here our problem is Prob>F is still not smaller than 0.05 so we can't conclude that they both can be a good variables for explaining gdpgrowth(we can't reject the hypothesis that both coefficients are equal to zero). Here coefficients represent how %1 change in variables effect dependent variable, holding other variables

constant. 0,01(%1) increase in gini results -0.47(-%0.47) change in gdpgrowth(both terms in percentage). Actually in U.S. for example, GDP seems rather positively correlated with gini, our result here is different from what we expected. So, we decided to add one more variable. But before that, we decided to see the effect of gini&gross both on the dependent variable, which we thought logical to see the result.

(3) We ran “regress gdpgrowth gini gross, robust” to see the results.

Coefficient of gini: -65.11987	std: 23.21359	p-value: 0.012	Prob>F: 0.0317
Coefficient of gross: 0.3765009	std: 0.1754487	p-value: 0.047	R-squared: 0.2811
Coefficient of constant: 24.13133	std: 8.72633	p-value: 0.013	
Corr(gini, gross): 0.4409			

The result is worth pointing that now both variables together seemed to explain growth in GDP, indicated by Prob>F smaller than 0.05. But now our p-values of both coefficients turned out to be insignificant(>0.05). Coefficient of gini is still negative and much more smaller than the latter result. Coefficient of gross is logical because we expect that %1 increase in gross saving(of GDP, by percentage) reflects a higher GDP level and higher GDP growth. Correlation of gini and gross is not very large so that we don't suspect of high multicollinearity here.

(4) We decided to see the regression of gini gross unemp together on gdpgrowth.

Coefficient of gini: -67.93544	std: 24.32729	p-value: 0.013	Prob>F: 0.0760
Coefficient of unemprate: -.1701277	std: 0.5775892	p-value: 0.772	R-squared: 0.2827
Coefficient of gross: .3714345	std: .1896709	p-value: 0.068	
Coefficient of constant: 27.1714	std: 12.71671	p-value: 0.048	
Corr(gross, unemp): -0.3318			

First of all, we saw that our Prob>F is not small enough to conclude that all independent variables are enough to see the overall effect on dependent variable. We can't reject the hypothesis that all coefficients are equal to zero. In addition to that, unemprate's and gross' p-values are also not small enough to reject that the coefficients are equal to zero, whereas gini's p-value is $0.0113 < 0.05$, statistically significant. Here, unemprate's coefficient's sign is still negative, that seemed understandable to us. Gini's coefficient is still negative and the smallest of all our regressions we had done so far.

Just a little change in both gini' and gross' coefficient and standard deviations in (4) comparing to (3) made us think that unemp hadn't caused any changes and it's better not to add unemp in our regressions anymore. It's coefficient is also not statistically significant.

(5) After excluding unemp, we decided to run “regress gdpgrowth gini gross popgrowth, robust”

Coefficient of gini: -116.9901	std: 34.61096	p-value: 0.004	Prob>F: 0.0235
Coefficient of gross: 0.0515739	std: 0.1721565	p-value: 0.768	R-squared: 0.3324
Coefficient of popgrowth: 11.19302	std: 6.286128	p-value: 0.094	
Coefficient of constant: 36.46784	std: 8.850409	p-value: 0.001	
Corr(popgrowth, gross): 36.46784	Corr(gini, popgrowth): 0.7737		

First of all, we should see how coefficient of gini differed here in comparison to (3). We thought that there may be omitted variable bias in (3) and with this equation(5) we eliminated it. Our gini's coefficient highly significant. Popgrowth's coefficient's sign is positive and corr(gini, popgrowth) is positive so the bias is positive, as we see here that the gini in (5) is very low comparing to (3). It sounded better to us, holding gross and popgrowth constant and to see gini's effect on average. Additionally, our R-squared is the highest one and Prob>F is the smallest one which is smaller than 0.05. They together seem to explain the change in gdpgrowth altogether. We can reject that all coefficient's are equal to zero, with 95 confidence.

But on the other hand, coefficient of gross and popgrowth seem insignificant by statistically, which is not a good thing for our regression. Also, coefficient of gross is very high. The other problem is corr(gross, popgrowth) is high. It doesn't cause perfect multicollinearity but it is enough for suspicion of imperfect multicollinearity, which causes variance of coefficient of gini to be very high. In comparison to (3), gini's coefficient standard deviation in (5) is very high. *From 23.21359 to 34.61096.* So it made us think once more what we could done about the regression.

(6) Then we decided to run "regress gdpgrowth gini popgrowth, robust". Maybe we could eliminate high multicollinearity and could satisfy the p-value and significant coefficients. But also we were not highly comfortable to eliminate such an important variable in terms of our thoughts.

Coefficient of gini: -81.46844	std: 35.93652	p-value: 0.035	Prob>F: 0.1016
Coefficient of popgrowth: 10.3971	std: 5.134953	p-value: 0.056	R-squared: 0.2538
Coefficient of constant: 23.99488	std: 8.755495	p-value: 0.013	

In this regression results reminded us of (2) which we added only unemprate in addition to gini in our regression. Significance result seems satisfying here but when we come to the Prob>F value which is larger than 0.05, resembling (2), we were not satisfied at all. Popgrowth's coefficient and standard deviation are not very different from (5). For one second, it caused us to think whether adding gross in (5) is a good idea or not because there is almost no change in popgrowth's coefficient and standard deviation. Also, standard deviation of gini is very high comparing to (3).

(7) To see what is our regression like when we add all variables here we ran "regress gdpgrowth gini popgrowth gross unemprate, robust"

Coefficient of gini: -115.4242	std: 35.3508	p-value: 0.005	Prob>F: 0.0209
--------------------------------	--------------	----------------	----------------

Coefficient of popgrowth:	11.36511	std: 6.628789	p-value: 0.107	R-squared: 0.4389
Coefficient of gross:	0.0508307	std: 0.1762049	p-value: 0.777	
Coefficient of unemprate:	0.1428082	std: 0.5184786	p-value: 0.787	
Coefficient of constant:	34.10563	std: 12.83513	p-value: 0.018	

The results seem very close to the results in (5), here in (7) we only have unemprate in addition to the variables in (5). So we were more sure that unemprate is not statistically significant and also doesn't even change our other variables' standard deviations or coefficients in a logical matter. R-squared seems the highest one here but it is not because this model is explaining more about the change in gdpgrowth, it's because when we add one more variable, R-square always increases, as we see in this regression. Yes, F value seems significant. But we exactly wouldn't prefer to run this regression to analyze the response of dependent variable. Additionally, the coefficients rather than gini is are not statistically significant.

As a result, we'd choose to run (3) or (5) to see the independent variables' effect on our dependent variable, especially to see the gini's effect on gdpgrowth. What we concluded is (3) is more preferable to (5) because of high standard deviation of gini and a suspicious multicollinearity in (5). On the other way, to run such regressions only with two variables may also not satisfying in order to see the real results and to reach an agreeable, logical causation. By software, we eliminated heteroskedasticity by robust term but we can't even observe or realize what is hid in the error term. Also, our sample is not large, which is another problem especially for variances of our coefficients.

A regression that is ran on U.S. data or some other countries, might show us very different results. We thought of adding many more variables such as industrial production, exports, imports, government expenditure but these all would cause reverse causal effect that our dependent variable would also affect the independent variables, which make them all bad control variables.

To summarize, according to (3), %1 increase in gini would cause %0.65 decrease in gdpgrowth, which means that a little increase in income inequality causes gdpgrowth to be lesser. One percent increase in gross, *gross savings as percentage of GDP*, would cause %0.37 increase in gdpgrowth.

IV. Bibliography