Table 1. Summary statistics: Income Inequality across Urban Areas. Average values from 2005 to 2009

	count	mean	sd	skewness	min	p5	p25	p50	p75	p95	max
Income share of the top 5%	952	19.95	2.26	0.73	14.40	16.60	18.50	19.90	21.30	23.60	31.40
Gini Index	953	44.02	3.19	0.44	35.00	39.00	42.00	44.00	46.00	50.00	56.00
Labor Force Participation Rate	953	62.01	6.55	-0.67	31.00	50.80	58.30	62.50	66.40	71.60	83.40
Unemployment Rate	953	43.47	7.05	0.51	20.50	32.60	38.60	43.20	47.50	55.30	72.60
Number of households	953	111470.61	362874.43	10.17	3911.00	8686.00	15128.00	28297.00	65746.00	434681.00	6750902.00
Households with Income in the past 12 months below poverty level	953	14361.29	43728.55	10.51	325.00	1129.00	2393.00	4408.00	9584.00	56535.00	825316.00
Household with public assistance income	953	2688.97	9589.24	11.86	14.00	148.00	326.00	657.00	1607.00	10344.00	187403.00
Total Population	953	300289.28	1020202.63	10.48	11100.00	22094.00	39467.00	73459.00	170049.00	1128813.00	18912644.00
Per capita income in the past 12 months (in 2009 inflation-adjusted dollars)	952	22332.35	4745.29	0.54	6022.00	15794.00	19580.50	21900.50	24791.00	30319.00	48394.00
Population 16 Years and Older	953	234638.29	795600.18	10.63	8353.00	17353.00	30983.00	58088.00	133546.00	909531.00	14959934.00
Population between 25 and 64 years old	953	159011.69	554451.49	10.58	4872.00	10920.00	19819.00	38129.00	85341.00	601916.00	10359750.00
Population with a High school diploma	953	43535.24	135853.56	10.69	973.00	3344.00	7013.00	12716.00	27526.00	167224.00	2647074.00
Population with some college degree or associate's degree	953	46728.91	146260.70	8.77	1193.00	3204.00	5965.00	11570.00	27447.00	184407.00	2376327.00
Population with a Bachelor's degree or higher	953	48180.99	198697.67	11.57	409.00	1773.00	3577.00	7397.00	21240.00	178000.00	3969791.00
Male	953	147772.51	499957.73	10.34	5786.00	10949.00	19601.00	36235.00	84263.00	551464.00	9147392.00
Female	953	152516.77	520357.17	10.62	5265.00	11164.00	19970.00	36990.00	85619.00	584966.00	9765252.00
White alone (population)	953	221425.87	654358.41	9.02	3458.00	16772.00	33648.00	61613.00	142470.00	779254.00	11474517.00
Black or African American alone (population)	953	37723.27	170123.90	11.38	0.00	85.00	706.00	3183.00	16122.00	150387.00	3326618.00
American Indian and Alaska Native alone (population)	953	2059.16	6617.85	7.06	0.00	33.00	136.00	359.00	1249.00	7600.00	87858.00
Asian alone (population)	953	13774.57	94605.77	14.45	0.00	56.00	236.00	616.00	2528.00	34647.00	1756296.00

Hispanic or Latino (population)	953	50692.23	287137.36	12.67	35.00	388.00	1151.00	3832.00	13740.00	140903.00	5614718.00
Married (population)	953	119077.77	387615.11	10.17	3666.00	9190.00	16719.00	30937.00	70210.00	436745.00	7136889.00
Not Married (population)	953	115560.52	408673.00	11.07	3908.00	7756.00	14337.00	27097.00	65028.00	446790.00	7823045.00
Households below poverty as % of total	953	15.81	6.95	2.48	4.17	8.28	11.50	14.33	18.17	28.13	60.27
Households with assistance income as % of total	953	2.44	1.21	1.79	0.13	1.05	1.63	2.21	2.92	4.79	10.77
Bachelors or higher as % of population	953	11.22	4.28	1.29	3.68	6.04	8.28	10.24	13.58	19.33	36.50
White as % of total population	953	82.33	14.21	-1.42	15.50	53.60	75.36	86.65	93.31	96.91	98.71
Black as % of total population	953	9.40	13.02	2.15	0.00	0.28	1.10	3.60	12.02	38.57	82.35
Hispanic as % of total population	953	10.95	17.92	3.12	0.17	0.92	1.84	4.17	10.56	48.67	99.87
Married as % of population	953	41.55	4.39	-0.51	24.61	33.71	38.94	41.65	44.63	47.93	54.67
Observations	953										

Note: Removed COLI (Cost of Living Index) due to the missing data and format issues.
Source: NHGIS (National Historic Geographic Information System), the U.S. Census Bureau.

Table 1. Summary statistics: Income Inequality across Urban Areas. Average values from 2015 to 2019

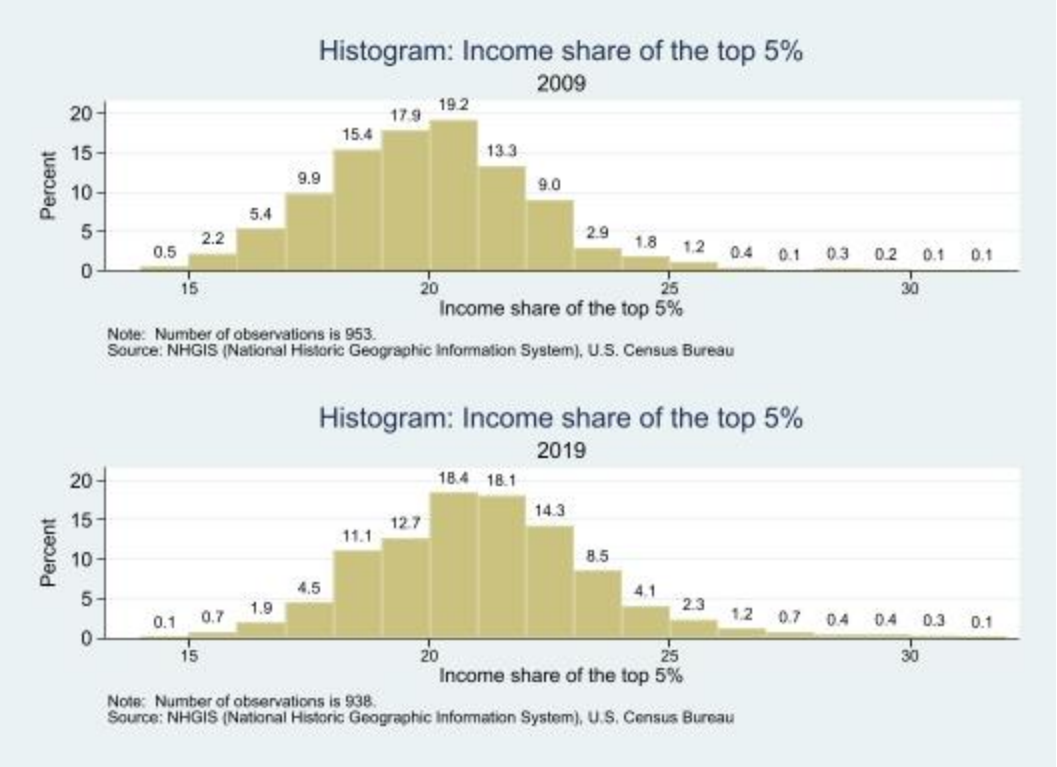
	count	mean	sd	skewness	min	p5	p25	p50	p75	p95	max
Income share of the top 5%	938	21.13	2.35	0.64	14.51	17.54	19.59	21.04	22.52	25.21	31.04
Gini Index	938	45.35	3.06	0.42	38.00	41.00	43.00	45.00	47.00	51.00	59.00
Labor Force Participation Rate	938	59.61	6.92	-0.72	22.50	47.90	55.60	60.15	64.30	70.10	79.10
Unemployment Rate	938	5.64	2.31	2.29	1.40	2.90	4.20	5.30	6.50	9.20	24.30
Number of households	938	122375.47	391641.89	9.51	3548.00	8322.00	15306.00	29376.50	71422.00	481217.00	6997001.00
Households with Income in the past 12 months below poverty level	938	15975.74	48927.95	10.18	425.00	1083.00	2425.00	4563.50	10712.00	61515.00	905007.00
Household with public assistance income	938	2961.67	10371.12	11.72	14.00	119.00	310.00	670.50	1810.00	12039.00	207019.00
Total Population	938	330282.28	1093044.10	9.66	12813.00	22164.00	39449.00	75436.00	182161.00	1257088.00	19294236.00
Per capita income in the past 12 months (in 2009 inflation-adjusted dollars)	938	28096.89	6162.14	0.72	7308.00	19198.00	24285.00	27466.00	31198.00	38725.00	60746.00
Population 16 Years and Older	938	264079.29	875321.50	9.79	9947.00	17230.00	31540.00	60833.50	152141.00	1005538.00	15599578.00
Population between 25 and 64 years old	938	224054.69	749014.43	9.86	6953.00	14443.00	26531.00	52073.50	126524.00	867371.00	13450844.00
Population with a High school diploma	938	50453.00	155152.00	10.20	1057.00	3681.00	7712.00	13990.00	32124.00	181592.00	2927576.00
Population with some college degree or associate's degree	938	18980.86	55549.99	8.27	303.00	1155.00	2407.00	4969.00	12029.00	72917.00	898607.00
Population with a Bachelor's degree or higher	938	73821.04	287993.58	10.51	907.00	2444.00	5145.00	11217.50	34512.00	267533.00	5430506.00
Male	938	162364.29	533976.96	9.56	6266.00	11148.00	20034.00	37316.50	91931.00	612116.00	9327459.00
Female	938	167917.99	559143.53	9.76	5874.00	10740.00	19855.00	37490.00	93488.00	643266.00	9966777.00
White alone (population)	938	236818.65	681403.91	8.20	4860.00	16342.00	33377.00	61931.00	150126.00	895445.00	11102463.00
Black or African American alone (population)	938	42774.55	184309.63	10.29	0.00	181.00	817.00	3327.00	16431.00	169054.00	3342737.00
American Indian and Alaska Native alone (population)	938	2411.51	7880.65	7.45	0.00	31.00	143.00	394.00	1505.00	9042.00	107728.00
Asian alone (population)	938	19003.03	121552.19	13.35	0.00	67.00	282.00	761.50	3538.00	48699.00	2166784.00
Hispanic or Latino (population)	938	64514.27	331073.78	11.51	14.00	620.00	1842.00	5690.00	19579.00	192970.00	5967160.00
Married (population)	938	128011.70	415678.21	9.51	4179.00	8573.00	15906.00	29539.50	73746.00	472671.00	7332129.00
Not Married (population)	938	136067.59	460514.75	10.04	4424.00	8332.00	15826.00	30077.50	75945.00	521251.00	8267449.00
Households	938	15.53	6.44	2.66	4.52	8.57	11.61	14.28	17.93	25.97	61.55

Households with assistance income as % of total	938	2.43	1.71	5.98	0.21	0.97	1.53	2.08	2.91	4.54	25.31
Bachelors or higher as % of population	938	16.22	5.99	1.05	4.56	8.71	11.92	14.94	19.58	27.28	47.99
White as % of total population	938	81.59	14.39	-1.50	14.39	52.48	75.33	86.38	92.23	96.23	97.77
Black as % of total population	938	9.28	12.70	2.22	0.00	0.51	1.38	3.77	11.52	37.33	77.62
Hispanic as % of total population	938	13.49	19.00	2.72	0.07	1.52	3.04	6.23	13.97	56.89	99.75
Married as % of population	938	39.81	4.26	-0.41	22.25	32.15	37.45	40.08	42.58	46.13	58.12
Observations	938										

Note: Removed COLI (Cost of Living Index) due to the missing data and format issues.
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau.

Based on the MSA-level dataset which captures the amount of income inequality across different MSAs for the period of years between 2005-2019, we summarized the number of observations, mean, standard deviation, skewness, minimum value, P5, P25, P50, P75, and maximum value for the 24 variables by using STATA. In order to examine the effects of the two determinants that contribute to income inequality, our group chose "Population with a Bachelor's degree or higher" as the main variable and the “Unemployment rate”, “White alone (population)”, “Black or African American alone (population)”, “Hispanic or Latino (population)”, “Married (population)”, “Household with public assistance income” and “Households with Income in the past 12 months below the poverty level” as the control variables. The correlation and degree of influence between the variables and determinants are further determined by comparing the data of the two periods, 2005-2009 and 2014-2019.

Illustrations:



We can observe that the overall inequality increased, with the shift of the distribution of the income share attributed to the top 5% earners to the larger values.

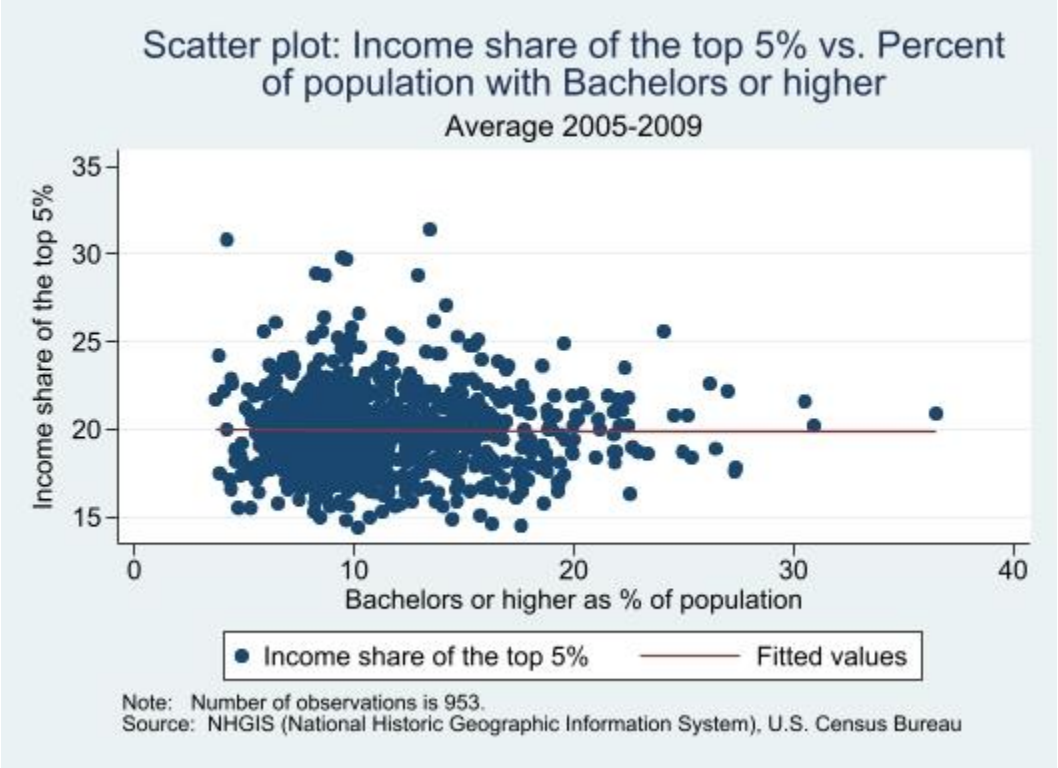
Table 1. Correlation between Income share of the top 5% and some of the potential main explanatory variables

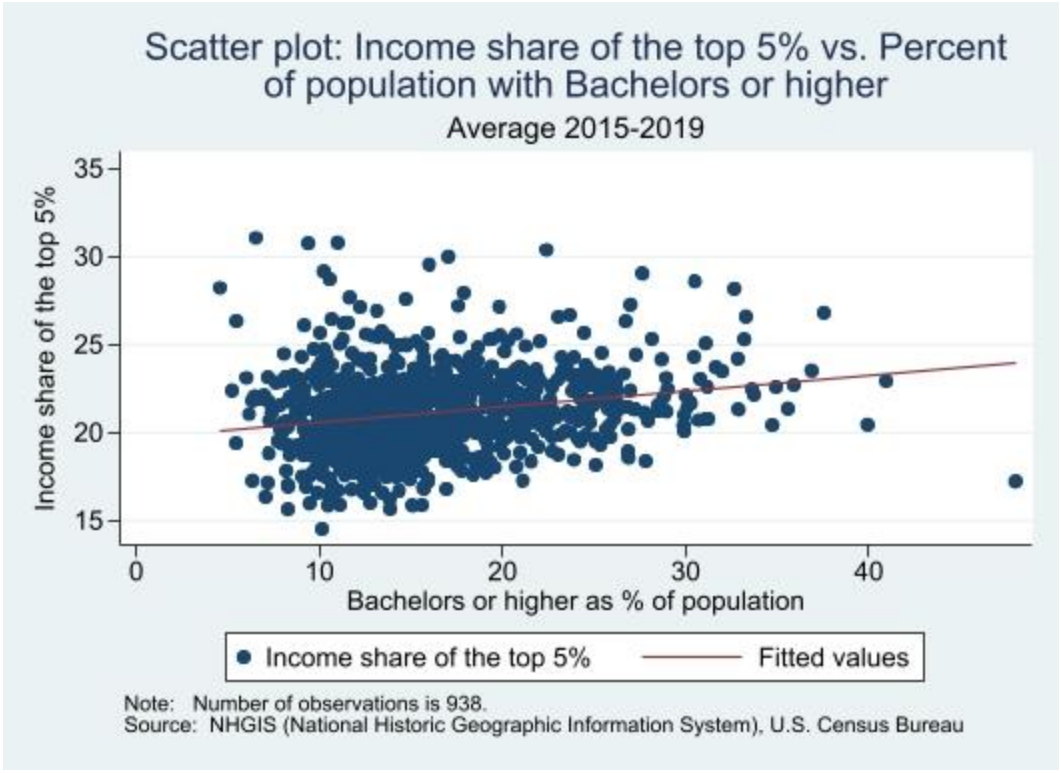
	Incomeshareofthetop5	degree	UnemploymentRate	whiteratio	blackratio	hispanicratio	married	hhassistance	hhpoverty
Incomeshareofthetop5	1								
degree	.2180533	1							
UnemploymentRate	-.1593038	-.3754472	1						
whiteratio	-.184524	.0248426	-.0518378	1					
blackratio	.1916859	-.1007414	.0666081	-.7907988	1				
hispanicratio	.0945114	-.104219	-.0047658	-.1438665	-.1507888	1			
married	-.2377916	.0098085	.1017929	.6403699	-.5347615	-.3223661	1		
hhassistance	-.0006529	-.1142723	.0047529	-.0714022	-.0966747	.2525937	-.1938158	1	
hhpoverty	.1875638	-.3281168	.0774774	-.3733168	.3450324	.4467398	-.6366918	.3198654	1
N	2819								

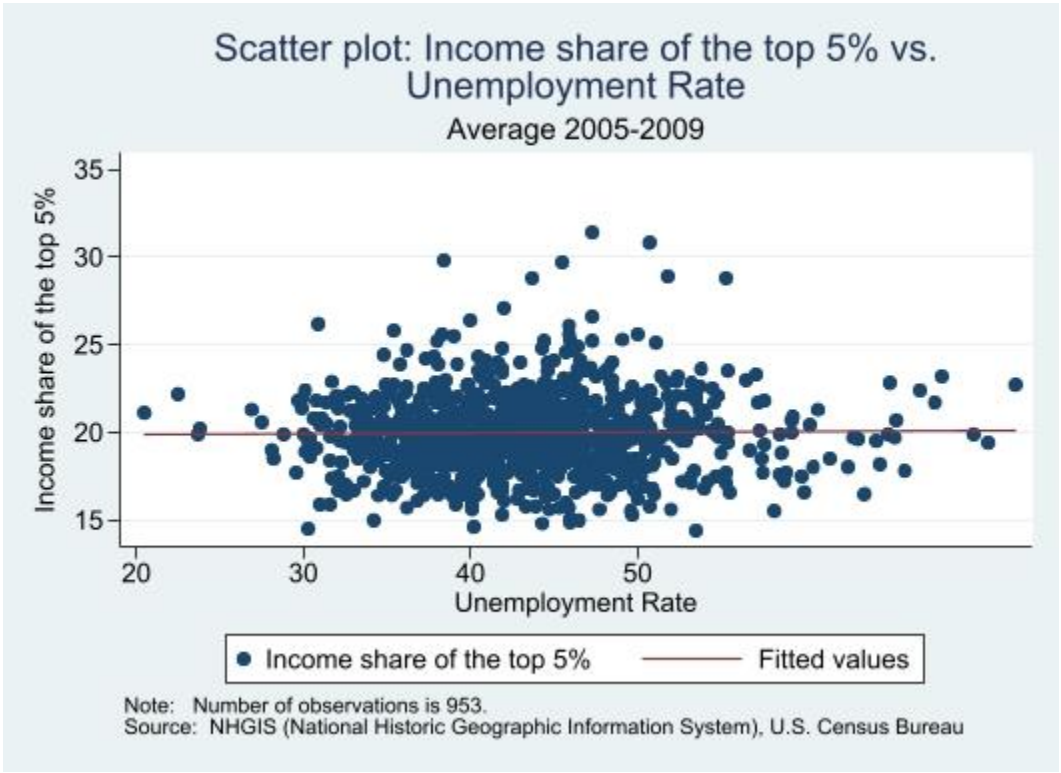
Note: .
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau.

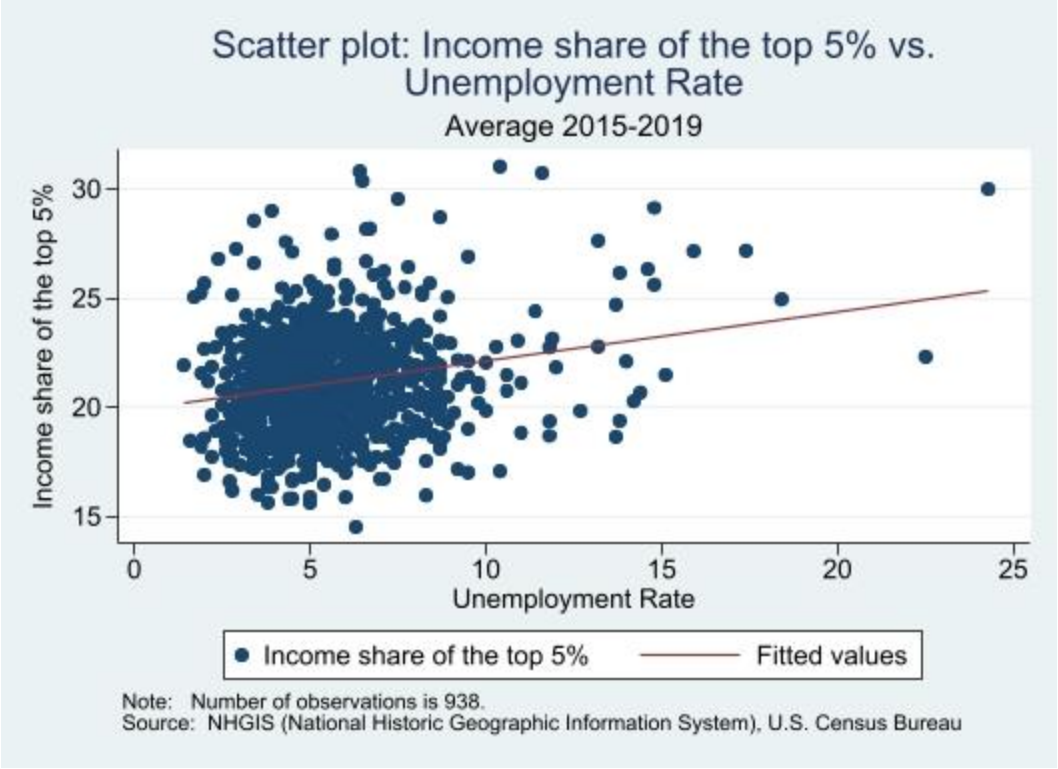
This correlation matrix shows potential explanatory variables that we will analyze. Our main independent variable is the percent of the population with Bahcelor's degree or higher (degree). Given the OVB conditions, we identify three variables of interest: Unemployment Rate, black ratio (percent of the total population that is black), and hhpoverty (percent of the total households that are below poverty rates.

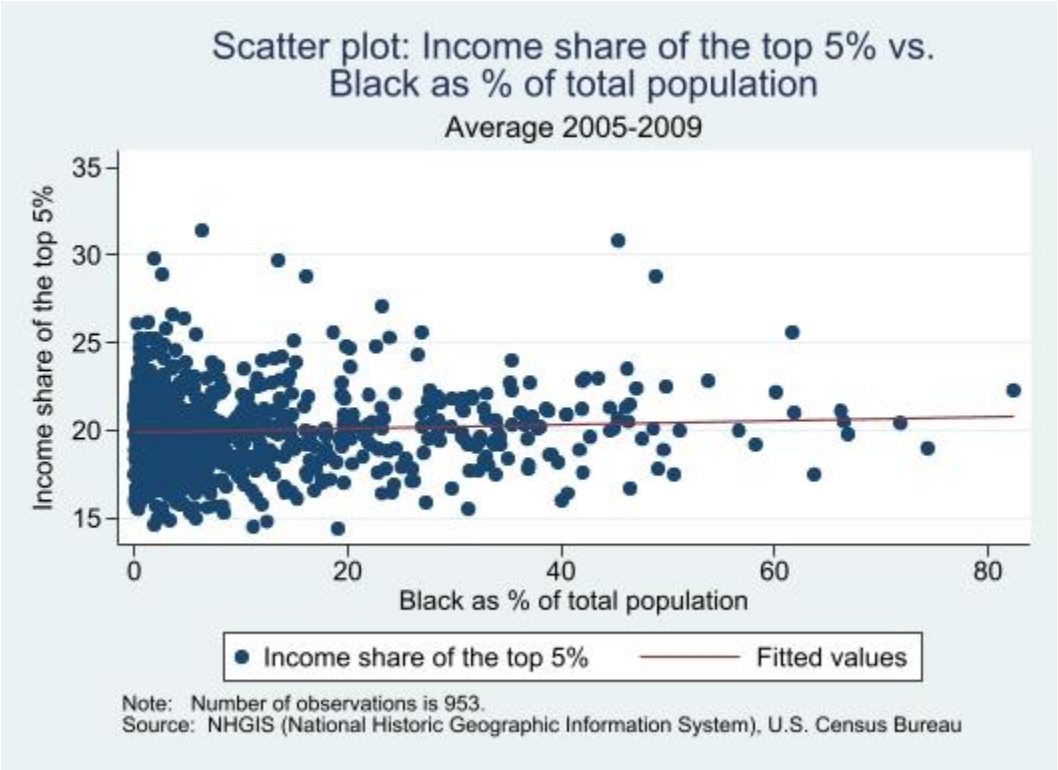
These variables are correlated with our dependent variable and the main X. The correlations, however, are only relatively strong, suggesting that there might be no significant relationship.

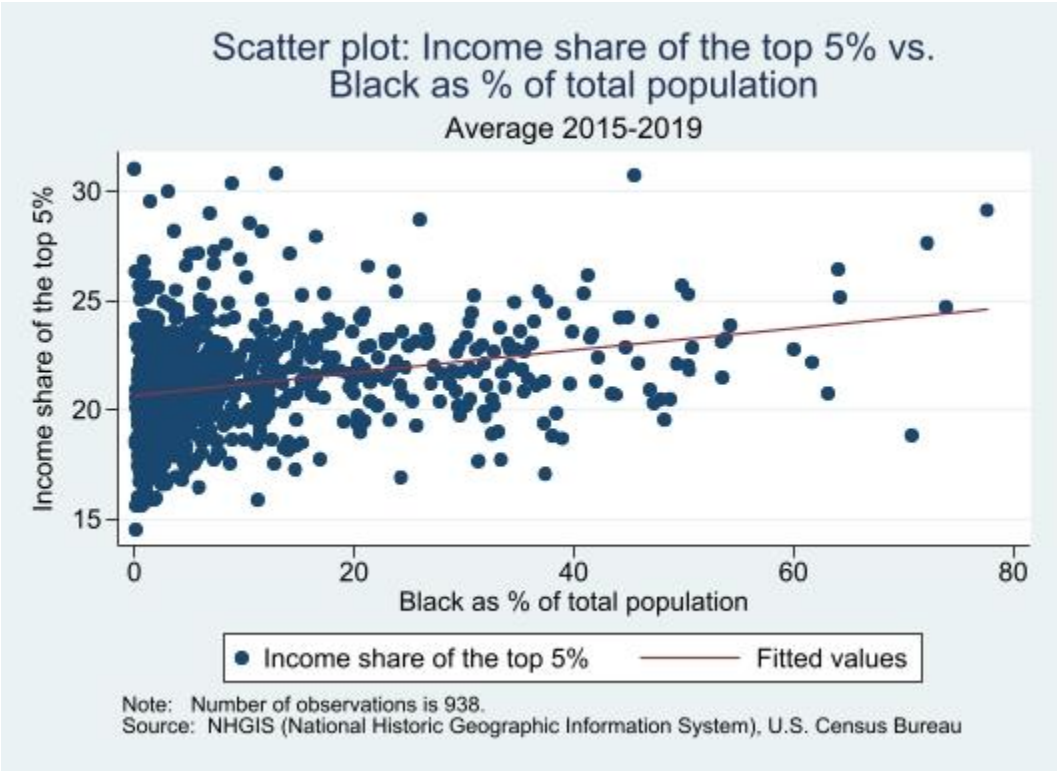



Scatter plot: Income share of the top 5% vs. Percent of population with Bachelors or higher
Average 2015-2019
Income share of the top 5%
35
30
25
20
15
0
10
20
30
40
Bachelors or higher as % of population
Income share of the top 5%
Fitted values
Note: Number of observations is 938.
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau


Scatter plot: Income share of the top 5% vs. Unemployment Rate
Average 2005-2009
Income share of the top 5%
35
30
25
20
15
20
30
40
50
Unemployment Rate
Income share of the top 5%
Fitted values
Note: Number of observations is 953.
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau


Scatter plot: Income share of the top 5% vs. Unemployment Rate
Average 2015-2019
Income share of the top 5%
30
25
20
15
0
5
10
15
20
25
Unemployment Rate
Income share of the top 5%
Fitted values
Note: Number of observations is 938.
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau


Scatter plot: Income share of the top 5% vs. Black as % of total population
Average 2005-2009
Income share of the top 5%
35
30
25
20
15
0
20
40
60
80
Black as % of total population
Income share of the top 5%
Fitted values
Note: Number of observations is 953.
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau



From the three pairs of scatter plot diagrams above, it can be found that there is almost no correlation between the three variables ("Population with a Bachelor's degree or higher", "Unemployment rate" and "Black or African American alone (population)") and the income share of the top 5% during the period of 2005-2009 (the line of Fitted values close to horizontal). However, it is clear to observe that there are positive growth relationships between these three variables and the income share of the top 5% during the period of 2015-2019.

To further check the relationship between the Income share of the top 5% and our dependent variable we perform regression analysis. Given that we have panel data, we first utilize Pooled OLS and Fixed Effects Regressions. Fixed Effects regression can help us control for potentially omitted variables that are absent from our dataset.

Table 1. Pooled OLS Regression results for Income share of the top 5%

	(1) Income share of the top 5%	(2) Income share of the top 5%	(3) Income share of the top 5%	(4) Income share of the top 5%
ln_degree	0.286***	0.262***	0.0195	-0.311***
	(0.0294)	(0.0296)	(0.0446)	(0.0923)
ln_unemploy		-0.301***	-0.367***	-0.445***
		(0.0456)	(0.0460)	(0.0465)
ln_blackratio			0.226***	0.138***
			(0.0292)	(0.0359)
ln_hhpoverty				0.554***
				(0.133)
Constant	17.83***	18.81***	19.40***	18.65***
	(0.287)	(0.336)	(0.342)	(0.391)
Observations	2819	2819	2817	2817
R^2	0.033	0.048	0.071	0.079
Adjusted R^2	0.032	0.048	0.070	0.078
F	94.38	82.33	85.61	75.64
rmse	2.273	2.255	2.221	2.212

Note 1: Robust standard errors are displayed in parenthesis.
Note 2: Y is in percentage points.
Significance levels: * p<0.10; ** p<0.05; *** p<0.01
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau.

In the first regression, we use a linear-log model, since our dependent variable is in percentages, it might be interesting how to decrease it by percentage points. The table illustrates the side-by-side comparison of different pooled-OLS outputs with statistical significance and precision measures. It is obvious that the addition of correlated variables from our selection does not produce meaningful improvement of statistical significance or precision, since rmse decreases while standard errors keep increasing. And even though p-values suggest the statistical significance of each variable, we see that the practical effects are rather small in all generated models. To have the largest effect out of this sample (0.554 percentage points increase) the population of the households below poverty must increase by 100%! The relationship, accordion to these models, is close to none.

Table 1. Fixed Effects Regression results for Income share of the top 5%

	(1) Income share of the top 5%	(2) Income share of the top 5%	(3) Income share of the top 5%	(4) Income share of the top 5%
ln_degree	2.480***	2.242***	2.266***	2.461***
	(0.230)	(0.398)	(0.404)	(0.395)
ln_unemploy		-0.0582	-0.0735	-0.0619
		(0.0871)	(0.0872)	(0.0855)
ln_blackratio			-0.175	-0.160
			(0.190)	(0.189)
ln_hhpoverty				-0.871**
				(0.402)
Constant	-2.907	-0.506	0.730	6.271
	(2.169)	(3.940)	(3.992)	(4.912)
Observations	2819	2819	2817	2817
R^2	0.080	0.081	0.081	0.084
Adjusted R^2	0.080	0.080	0.080	0.083
F	116.7	59.26	39.34	31.97
rmse	1.572	1.572	1.569	1.567

Note 1: Robust standard errors are displayed in parenthesis.
Note 2: Y is in percentage points.
Significance levels: * p<0.10; ** p<0.05; *** p<0.01
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau.

These are other linear-log models, however, a theory is derived using fixed effects regression. The effect is still measured in percentage points. Here we observe a similar situation, where rmse gets worse and standard errors rise with almost every addition. Additionally, adjusted R squared basically stagnates. However, these models might be an improvement, since the effect of degree is now larger, while other variables lost their significance (bar hhpoverty). Nevertheless, the relationship between the population of university graduates and the income share of the top 5% is still very irrelevant. For a 2.4 percentage point increase in the income share attributed to the top 5%, there must be an increase in the population of university graduates by 100%. Too impractical of an effect, even though it might be statistically significant.

Table 1. Fixed Effects Regression results for logged Income share of the top 5%

	(1) ln_incshare	(2) ln_incshare	(3) ln_incshare
ln_degree	0.121***	0.129***	0.117***
	(0.0111)	(0.0113)	(0.0186)
ln_hhpoverty		-0.0411**	-0.0395**
		(0.0196)	(0.0195)
ln_unemploy			-0.00367
			(0.00411)
ln_blackratio			-0.00732
			(0.00946)
Constant	1.868***	2.154***	2.319***
	(0.105)	(0.176)	(0.231)
Observations	2819	2819	2817
R^2	0.082	0.085	0.086
Adjusted R^2	0.082	0.085	0.084
F	120.2	64.95	33.22
rmse	0.0759	0.0758	0.0758

Note 1: Robust standard errors are displayed in parenthesis.
Significance levels: * p<0.10; ** p<0.05; *** p<0.01
Source: NHGIS (National Historic Geographic Information System), U.S. Census Bureau.

Here we use logged variables to observe the percentage changes in the dependent variable with the changes in our main variable of interest, degree. Again, the rmse, standard errors, and adjusted R squared do not indicate an improvement with the addition of correlated variables from the sample. The models demonstrate that for every one percent increase in the population of university graduates, the income share of the top 5% change only by 0.12%. This means an increase of the graduates by 100% would yield only a 12% increase in the share of income. Either way, this is a big leap for such a relatively small change.

However, if we take into account the change in the graduates' population over time, this change can have some meaning. Assuming that it reflects a causal relationship, over time this effect can get more tangible and acute. For example, the number of people in the US with educational attainment of a Bachelor's or higher was 37.9% in 2021, which is almost a 100% increase from only 23% back in 1995, less than 30 years ago. We should take into account, though, that these models suffer from serious OVB since we did not include many other important determinants of inequality. Thus, the model might have absorbed the overall trend of increase in inequality. So this small finding should not be used for policy-making.