Econometric Analysis of Slum Population

This paper provides an analysis of key variables affecting the proportion of the population living in slums across various countries, using a combination of descriptive statistics, histograms, and econometric regressions. The analysis focuses on variables such as the proportion of refugees in a country, GDP per capita, income inequality, and urbanization, and examines their relationship with slum prevalence in urban areas. By creating a new variable, "percent_refugees," and exploring the results of both fixed effects and pooled OLS regressions, this paper seeks to shed light on the complex socio-economic factors influencing slum conditions.

Table of Descriptive Statistics

```
. tabstat *, stats(mean sd sk min p5 p25 p50 p75 p95 max) ///
>            column(statistics)
```

variable	mean	sd	skewness	min	p5	p25	p50	p75	p95	max
year	2001.031	6.751508	-.4138176	1990	1990	1995	2005	2007	2009	2009
proportion~l	53.20677	22.88394	-.0591393	3.3	15.8	34.1	54.65	70.1	91.3	98.9
gdp_per_ca~s	4108.207	3577.296	1.290923	434.0814	700.7804	1464.282	2694.07	5769.762	11588.17	16633.6
gdpgrowth	4.329275	4.740429	1.752042	-6.9	-3.1	2.4	4.5	6.2	9.8	36.1
unemployme~e	8.231973	6.116726	1.877636	.6	1.5	4.1	7.4	9.5	20.2	37.3
refugeesnu~r	160138.3	405815.1	4.331842	1	103	1748.5	17858	123972	878198	3255975
totalpopul~n	1.35e+08	2.93e+08	3.149707	2912824	4659458	1.27e+07	3.26e+07	1.11e+08	1.13e+09	1.32e+09
percenturban	40.23983	19.63303	.5309321	2.5	13.4	26.8	36.8	53.5	76.9	90.8
urbangrowth	3.745251	2.198081	3.108761	-.8	1.3	2.5	3.5	4.4	6.6	18.7
popdensity	109.776	174.8895	3.906966	1.4	4	20.9	54.4	126.7	357.9	1190.8
incomeshar~e	44.74868	10.22202	-.5619352	6.8	29.8	38	45.7	51.7	59.6	63.1
gini	5.571053	3.313279	3.452765	.8	2	3.4	5.05	7.05	9.3	26.7
incomesha~10	35.39737	7.733149	-.3490631	7.1	25	29.4	35.65	41.05	47.8	51.7
urbanpover~e	27.45714	14.29722	.1910776	1	9.1	13.2	28.6	36.7	51.1	61.5
phones	26.57857	28.95012	1.18727	.1	.2	2.5	15.7	41.35	87.3	131.1
internet	4.27625	8.464209	4.296431	.1	.1	.3	.75	5.25	18.95	69.4
healthcare~o	6.68069	8.997422	6.592009	.8	3.2	4.3	5.2	6.5	9.2	80.2
infantmort~e	45160.93	491046.5	10.83256	12.9	17.4	35.6	60	88.25	121.8	5800000
HDI	.5264151	.1304219	-.1358229	.2	.3	.4	.5	.6	.7	.8
government~s	-.6025676	.5743248	-.4614159	-2.34	-1.66	-.935	-.555	-.175	.23	.88
politicals~y	-.7546154	.8577457	-.4689868	-3.32	-2.3	-1.3	-.61	-.11	.47	1.02
con_code	30.42061	17.30078	.0012809	1	3	15	30	45	57	60
percent_re~s	.2531011	.5210095	3.35101	9.79e-06	.0001364	.0035232	.0238644	.2931734	1.240213	2.930911

Analyzing the Results of the Table of Descriptive Statistics

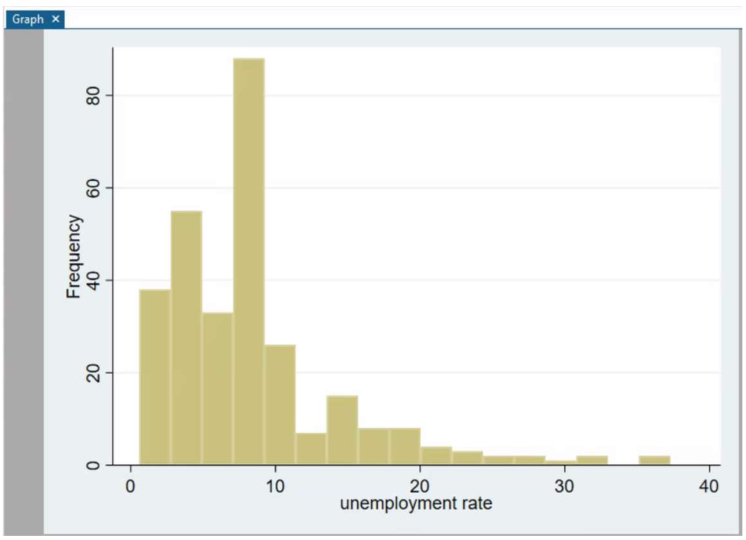
In conducting the econometric analysis, I created the variable percent_refugees by dividing the number of refugees in each country by its total population. This allowed for a clearer view of the proportion of refugees relative to the overall population. Additionally, several of the variables (GDP, Gini, and the number of refugees) showed strong rightward skewness. This indicates that a small number of countries in the dataset have extremely high values, while most countries exhibit lower values (we can see this in the histograms as well). This skewness is an important consideration in our analysis, as it suggests that the data might not follow a normal distribution, and we should be cautious about generalizing based on mean values alone.

One particularly interesting result was the data for income share of the top 10%, which had an average of 35% with a slight skew to the left. This percentage is lower than what I expected, especially when compared to wealthier nations like the United States, where the top

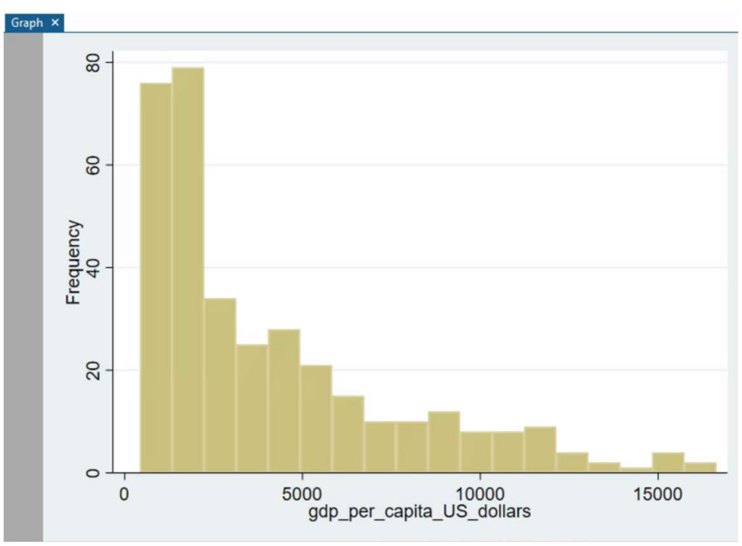
10% holds a significantly larger share of income. However, when looking at the countries included in this dataset as well as GDP (average GDP per capita of just 4,108 USD) and internet/phone access, it is clear that we are dealing with underdeveloped countries.

The mean proportion of the urban population living in slums was 53.2%, with a standard deviation of approximately 23 percentage points. This result was unexpectedly high and provides valuable insight into the living conditions within these countries. The high standard deviation suggests significant variation in slum prevalence across countries. Given the strong relationship between poverty, inequality, and urban slum conditions, understanding the impact of economic factors like GDP, income distribution, and access to services will be critical in analyzing the factors contributing to slum growth and poverty in urban settings.

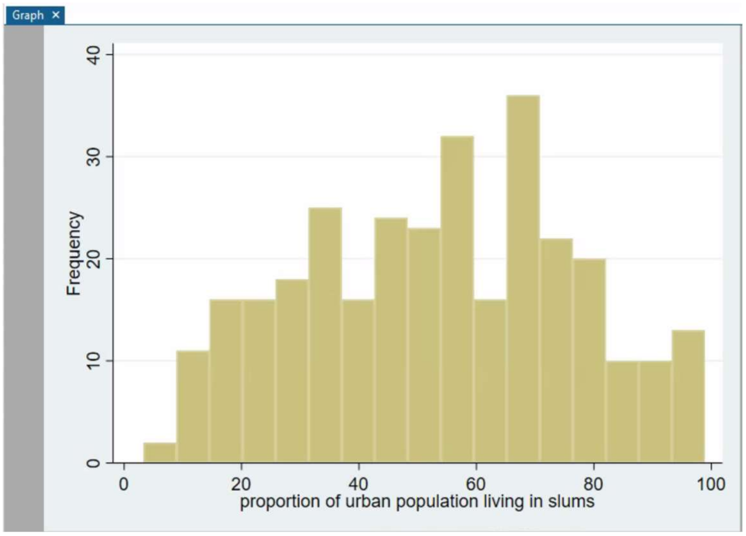
Histogram of Unemployment Rate



Histogram of GDP Per Capita (USD)



Histogram of Prop. Of Urban Pop. Living in Slums



Analyzing Histograms:

The first histogram, representing the frequency of data points for unemployment rates, shows a peak just under 10%, indicating that most data points cluster around this unemployment rate. There is a higher concentration of values to the left of this peak, suggesting that a large proportion of observations are associated with lower unemployment rates, though there are a few data points stretching towards much higher unemployment rates, reaching up to around 37%. This agrees with the high skew we saw in the table of descriptive statistics. The GDP per capita histogram peaks at the far left, representing a majority of countries with relatively low GDP per capita, and the number of countries steadily decreases as GDP per capita approaches $15,000. This distribution highlights the disparity in global wealth, with fewer countries achieving higher levels of economic output. Lastly, the histogram for the proportion of the urban population living in slums shows a peak around 60%, with a roughly normal distribution around this central value. This suggests that the majority of countries have a significant portion of their urban populations living in slums, but the distribution indicates a range of experiences, with some countries experiencing both higher and lower slum populations.

Table 1. Pooled OLS Regression results for Slum Population Panel Data

	(1) proportion~l	(2) proportion~l	(3) proportion~l	(4) proportion~l	(5) proportion~l
gdp_per_capita_US_dollars	-0.00433***	-0.00406***	-0.00374***	-0.00138	-0.00172*
	(0.000222)	(0.000230)	(0.000647)	(0.000924)	(0.000915)
year=1990	0	0	0		
	(.)	(.)	(.)		
year=1995	-4.088	2.435	6.834	0	
	(3.109)	(2.338)	(4.377)	(.)	
year=2000	-6.371**	-1.038	0.690	-0.747	0
	(3.085)	(2.476)	(4.332)	(4.695)	(.)
year=2005	-9.329***	-3.894	-1.794	-9.204	-12.23**
	(3.101)	(2.589)	(4.851)	(5.412)	(4.402)
year=2007	-8.249***	-3.145	-0.353	-7.018	-14.74*
	(3.050)	(2.588)	(5.214)	(6.926)	(8.067)
year=2009	-8.085***	-2.842			
	(3.033)	(2.413)			
unemployment rate		-0.355***	-0.197	-0.117	0.0688
		(0.131)	(0.306)	(0.350)	(0.427)
percent urban			0.0482	0.356**	0.759*
			(0.108)	(0.167)	(0.405)
percent_refugees			12.12***	30.76*	30.96*
			(2.975)	(15.91)	(16.00)
urbangrowth				3.353*	1.162
				(1.631)	(1.527)
popdensity				0.0366***	0.0334***
				(0.00953)	(0.00745)
income share of richest 10%				-0.438	-0.0324
				(0.473)	(0.509)
infant mortality rate				0.0233	-0.126
				(0.180)	(0.363)
HDI					-128.7
					(105.0)
government effectiveness					-15.60
					(10.09)
political stability					3.531
					(2.606)
Constant	77.26***	73.65***	61.92***	30.55*	86.63
	(2.310)	(2.055)	(5.671)	(17.13)	(72.40)
Observations	303	257	100	31	27
R-squared	0.537	0.545	0.560	0.804	0.899
Adjusted R-squared	0.527	0.532	0.521	0.691	0.797
F	68.78	.	.	29.75	34.49
rmse	15.56	15.27	16.13	9.658	8.317

Note 1: Robust standard errors are displayed in parenthesis.
Note 2: 359 Observations
Significance levels: * p<0.10; ** p<0.05; *** p<0.01
Sources: WHO, World Bank, ILO, United Nations

Table 1. Fixed Effect Regression results for Slum Population Panel Data

	(1) proportion~l	(2) proportion~l	(3) proportion~l	(4) proportion~l	(5) proportion~l
gdp_per_capita_US_dollars	-0.00146	-0.00175*	-0.00261	0.0000857	-0.000611
	(0.000883)	(0.000926)	(0.00183)	(0.000124)	(.)
year=1990	0	0	0		
	(.)	(.)	(.)		
year=1995	-4.769***	-37.41***	-41.92***	0	
	(1.294)	(3.430)	(5.604)	(.)	
year=2000	-8.811***	-41.76***	-46.68***	3.138***	0
	(1.709)	(3.367)	(4.799)	(0.505)	(.)
year=2005	-12.15***	-45.07***	-47.86***	-4.164***	0
	(2.017)	(3.036)	(3.451)	(0.967)	(.)
year=2007	-12.27***	-45.22***	-47.13***	-4.902***	0
	(2.225)	(3.031)	(3.656)	(1.185)	(.)
year=2009	-13.03***	-46.02***			
	(2.366)	(2.849)			
unemployment rate		-0.258	0.00277	0.341***	1.360
		(0.268)	(0.364)	(0.0870)	(.)
percent urban			-0.406	1.885***	19.05
			(0.591)	(0.317)	(.)
percent_refugees			-1.554	108.1***	-292.0
			(2.097)	(4.196)	(.)
urbangrowth				-21.53***	144.8
popdensity				0.328***	-1.875
				(0.0316)	(.)
income share of richest 10%				-0.213	-3.031
				(0.160)	(.)
infant mortality rate				1.649***	-0.674
				(0.0877)	(.)
HDI					0
					(.)
government effectiveness					-12.58
					(.)
political stability					0
					(.)
Constant	67.90***	104.2***	125.6***	-128.1***	-995.7
	(3.231)	(7.295)	(30.54)	(16.94)	(.)
Observations	303	257	100	31	27
R-squared	0.477	0.513	0.617	0.993	1.000
Adjusted R-squared	0.467	0.499	0.583	0.988	1.000
F	13.56	.	.	.	.
rmse	5.527	4.402	4.791	0.337	0

Note 1: Robust standard errors are displayed in parenthesis.
Note 2: 359 Observations
Significance levels: * p<0.10; ** p<0.05; *** p<0.01
Sources: WHO, World Bank, ILO, United Nations

Analyzing Regressions

When comparing the results from a fixed effects regression and a pooled OLS regression on panel data used to predict the proportion of the population living in slums, several key differences and insights emerge. For the fixed effects regression, the most surprising finding is the statistical insignificance of the final regression. This suggests that when controlling for individual unit characteristics over time, the model is less able to identify significant relationships between the predictors and the outcome. However, the most impactful regressors in the fixed effects model are the year dummies, which account for time-specific factors that affect all units equally. This result implies that year-to-year variations, such as global economic trends or policy changes, have a substantial impact on the proportion of the population living in slums. The inclusion of year dummies allows the fixed effects model to isolate the variation across units and capture the broader temporal trends that could otherwise confound the results.

A particularly interesting result from the fixed effects regression is the coefficient on percent_refugee, which is statistically significant at 108.1. This suggests that a 100 percentage point increase in the proportion of refugees corresponds to a 108 percentage point increase in the proportion of the population living in slums. While this effect seems unrealistic, it can be interpreted more reasonably as a 1 percentage point increase in the refugee population leading to a 1.08 percentage point increase in the slum population. This coefficient, though large, highlights the potential impact that an influx of refugees may have on urban housing and infrastructure, especially in areas with pre-existing challenges related to slum development.

In contrast, the pooled OLS regression results show a different pattern. Although year dummies are also included in the pooled OLS regressions, the statistical significance of the results diminishes as more regressors are added, particularly in the final model with 17 regressors. This is likely due to overfitting, as evidenced by the relatively low number of observations (only 27) and a high R-squared value of 0.899, which indicates that the model is explaining much of the variation but may be fitting noise rather than true underlying patterns. The small sample size and the large number of regressors also suggest that the model is overly complex, leading to a loss of statistical significance for many variables.

Another issue contributing to the weakened results in the pooled OLS model is multicollinearity between several independent variables. For instance, percent urban and population density are likely highly correlated, as more urbanized areas tend to have higher population densities. Similarly, the unemployment rate and GDP per capita are likely correlated, as poor economies usually have higher unemployment. The multicollinearity between these variables may explain the lack of a statistically significant effect for GDP per capita in regressions 4 and 5. This issue arises because the variables are not providing unique information to the model, which can distort the estimates and inflate the standard errors, further reducing the likelihood of finding significant effects.

In summary, the fixed effects regression provides a more reliable framework for identifying significant relationships between predictors and the proportion of the population living in slums by controlling for unit-specific and time-specific effects. On the other hand, the pooled OLS regression, with its potential issues of overfitting and multicollinearity, struggles to produce meaningful results, especially as more variables are added to the model. The comparison highlights the importance of choosing the appropriate regression technique based on the structure of the data and the specific challenges inherent in panel data analysis.