DETERMINANTS OF GDP, INEQUALITY AND THE RISK OF POVERTY ON A REGIONAL LEVEL

(Panel Data Econometrics Study Using Fixed and Random Effects)

Prepared by: BOZHIDAR KARANOVSKY, University of St.Gallen, MiQEF
intern in: Economic and Financial Policy Directorate,
Macroeconomic Analyses and Forecasts Division

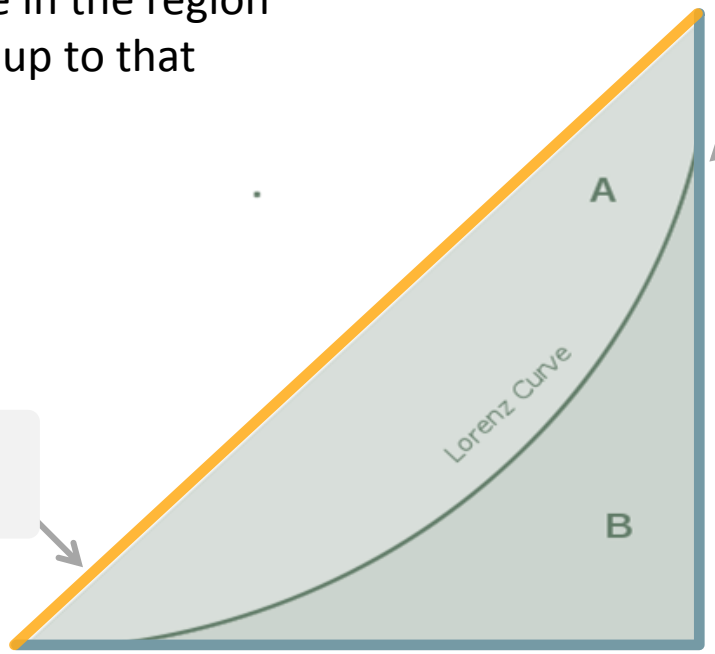
August 2014, Ministry of Finance of the Republic of Bulgaria

Inequality has been a hot topic lately…

GINI Index = A/(A+B)

X : households sorted by income

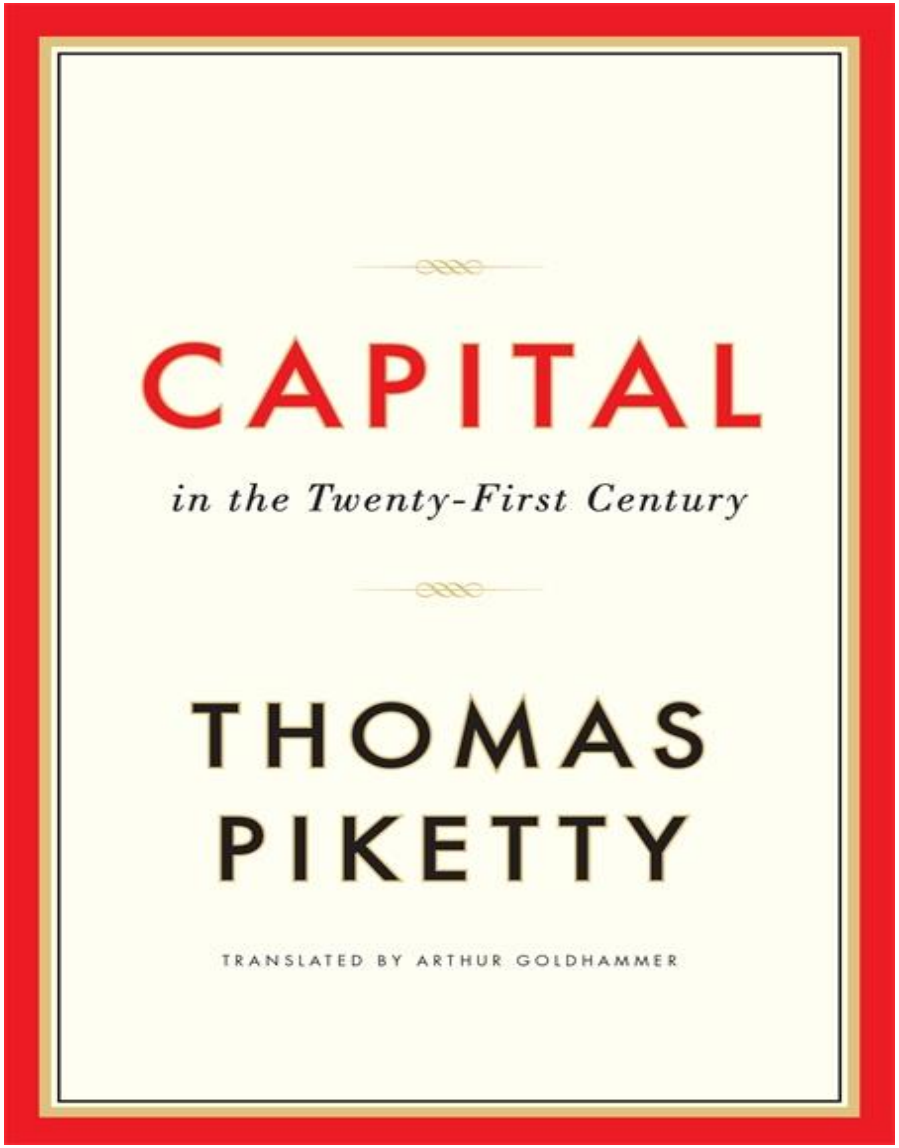
Y: cumulative income in the region going to households up to that percentile



Two perpendicular lines : perfect inequality (the household at the 100th percentile collects the whole income

Linear : Perfect equality

Inequality is natural.
But how much is too much?



#1 New York Times Bestseller

But really just repackaged old ideas…

Argumentation

- “There is a widespread agreement that income disparities across European regions have narrowed over time, but reduction of income disparities across regions cannot be equated with reduction of disparities within regions. That is, a region with high GDP per capita may have substantial pockets of poverty, and a region with low GDP per capita may have some areas of prosperity. The directives of the European Commission implicitly assume that the funding received by a region will be converted not only to greater prosperity on average, but will also reduce the existing disparities in the region. Resources awarded to a region whose average income level is low may simply result in additional well paid jobs for the narrow upper-middle class and, ultimately, in a greater inequality.” (Longford, Pittal et al., 2010)

- Much talk about “’feudalization” of regions by local power brokers. What drives GDP growth, inequality and poverty on a regional level?

- Key regressor whose effect on the 3 variables I am most interested in: investment.

- What are the correlations between these three variables and other important regional statistics? Can they be explained via a causal relationship? Most important – policy implications?

Sources

- *National Statistical Institute (www.nsi.bg)*
- *Eurostat (http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home)*
- *"Regional Profiles: Indicators of Development" Study by the Institute for Market Economy, 2013 (www.regionalprofiles.bg)*
- *Lechner, Michael. "Econometrics". University of St. Gallen, Lecture Notes, 2013*
- *Wooldridge, Jeffrey. "Introductory Econometrics. A Modern Approach". 4th Ed: South-Western, 2009*
- *Longford, Nicholas and Pittau, Maria Grazia and Zelli Roberto and Massari Riccardo. "Measures of Poverty and Inequality in the Countries and Regions in the EU". ECINEQ: Society for the Study of Economic Inequality. Working Paper Series 2010-182.*
- *Help from Ms. Albena Nikolova in particular*

Note: The interpretations and policy recommendations of the results of the study reflect only and exclusively the opinions of the author and are not necessarily indicative of the stances of any other institution, including the Ministry of Finance.

Problems…

- Short time series on GINI index, Income ratio, at risk of poverty rate on the regional level (2007-2011); virtually no reliable statistics on quality of life
- Public data on utilization of EU operational funds (per capita) only for 2011-2012 on regional level => their impact on inequality and poverty levels?
- Lack of price deflators on a regional level
- Less rich statistics on regions NUTS3 in general than national or NUTS2 level => possibility of confounders in error term.
- Possible Solution: Panel Data Fixed and Random effects

Structure of the Data

	Region	Year	GDP.pc	At.risk.of.Poverty.Rate	Income.Ratio	GINI	FDI.non.financial.enterprises.pc	Expenditures.for.Acquisitions.of.Fixed.Tangible.Assets.pc	Unemployment.Rate	Employment.Rat
1	Bulgaria	2007	7379	40.0	6.5	35.9	401.414082	3583.2125	6.9	49.0
2	Bulgaria	2008	9090	38.8	5.9	33.4	463.928103	3859.7225	5.6	50.8
3	Bulgaria	2009	9007	40.8	5.9	33.2	250.922683	2804.7347	6.8	49.4
4	Bulgaria	2010	9359	41.7	6.5	35.0	240.219345	2152.5945	10.2	46.7
5	Bulgaria	2011	10248	41.8	6.1	33.6	10.335148	2437.7835	11.3	45.6
6	Благоевград	2007	5175	27.4	3.7	22.3	103.348881	1729.9073	2.3	56.4
7	Благоевград	2008	6219	29.2	3.2	19.7	246.295636	2129.2555	1.8	57.2
8	Благоевград	2009	6032	34.0	3.0	20.4	115.444117	1319.3790	3.4	55.4
9	Благоевград	2010	5982	33.3	4.3	28.3	1.329385	1049.3213	5.8	53.6
10	Благоевград	2011	6682	29.9	3.5	23.3	26.713091	1228.4547	8.3	53.8
11	Бургас	2007	6876	35.9	6.9	36.0	140.535177	4053.4118	4.1	47.9
12	Бургас	2008	8609	37.2	6.0	32.5	-256.830901	4395.4189	3.4	49.1
13	Бургас	2009	8064	34.5	6.3	33.6	398.671812	3031.9726	3.9	48.3
14	Бургас	2010	8082	35.8	6.1	34.3	1862.560014	2039.0933	9.6	44.8
15	Бургас	2011	9277	34.9	5.5	31.1	513.873394	2378.5870	12.6	44.5
16	Варна	2007	8336	35.1	4.8	30.2	487.563456	4934.4570	6.8	53.2
17	Варна	2008	10227	43.6	5.7	32.2	898.208834	4500.5992	4.3	54.8
18	Варна	2009	9613	39.3	5.2	29.9	136.001637	3020.1029	4.3	52.9
19	Варна	2010	9595	42.4	5.8	31.8	-240.819212	2406.1851	8.2	49.6
20	Варна	2011	10270	42.9	5.1	28.7	-129.729610	2299.1104	10.3	46.5
21	Велико Търново	2007	4920	49.3	7.8	35.0	74.434039	1701.2897	6.5	44.6
22	Велико Търново	2008	5682	38.4	6.2	32.0	23.587555	1667.1161	7.8	44.6
23	Велико Търново	2009	5716	49.4	6.1	31.5	-56.205442	1026.7563	9.0	42.6
24	Велико Търново	2010	6135	44.6	8.9	37.6	-3.155591	979.8864	13.1	38.9

Variables Collected (1)

Variable	Description and Interpretation	Unit
GDP pc	Gross Domestic Product per capita. Measures the standard of living and the strength of the economy in the district.	BGN per capita
At Risk of Poverty Rate	The relative share of people living below the district's poverty line, which is defined as 60 percent of the regional median equivalent disposable income. This indicator was chosen over "relative share of population living in material deprivation". **Calculated before social transfers and pensions.**	%
Income Ratio	A measure of inequality. Ratio between the cumulative incomes of the top 20% and the bottom 20% of the households in a region.	%
GINI	Index for inequality. 0 signifies perfect equality (all persons having the same income), 1 signifies perfect inequality (one person receiving the whole income and all the others receiving zero).	%
FDI in Non-Financial Enterprises per capita	Annual inflow (if positive) or outflow / disinvestment (if negative) of Foreign Direct Investments in non-financial enterprises per capita to the district. It shows how attractive the region is to foreign investors. More FDI fosters economic growth, and theoretically should create jobs and therefore reduce poverty and inequality. But does the second part of this statement hold true?	BGN per capita
Expenditures for Acquisitions Of Fixed Tangible Assets per capita	The level of expenditures for acquisition of fixed tangible assets (FTA) per capita in the district. This reflects the level of investment in a district and the expectations by businesses for the future. It also reflects how much is invested in productive activities and availability of credit. Higher investment should lead to more employment which should reduce inequality, reduce poverty and raise GDP.	BGN per capita
Unemployment Rate	Annual average of the unemployment rate of the population in the district above the age of 15. Equals unemployed/labor force. Should be positively correlated with poverty and negatively with GDP.	%

Variables Collected (2)

Variable	Description and Interpretation	Unit
Employment Rate	Annual average of the population aged 15+ in the district. Calculated as employed/population aged 15+. It should reduce inequality and poverty and raise GDP.	%
Non-Financial Companies per 1000 people	The number of non-financial companies per 1000 people in the district. Used for proxy of entrepreneurship, which theoretically should foster GDP, investment and growth and reduce poverty.	Number of businesses / 1,000 people of population
Share of up to Lower Secondary Education	Does not include people who besides lower secondary education have completed secondary or tertiary education.	%
Share of Secondary Education	Does not include people who besides secondary education, have completed tertiary education.	%
Share of Tertiary Education	Share of the population who have completed tertiary education.	%
Population per General Practitioner	Indicator of the availability of the health services, and more specifically, the availability of medical staff relative to the population.	Population / number of general practitioners
Road Network Density	The total length of highways and roads (first, second and third class) divided by the total area of the region. Streets in urban areas are excluded! That is Sofia (capital) has a value of 0. Since this biases results, this variable is excluded in the poverty regression. Better infrastructure and easier transport of passengers and goods fosters growth, reduces costs and therefore should reduce poverty and inequality.	Length of the road network km / 100 sq. km. of area

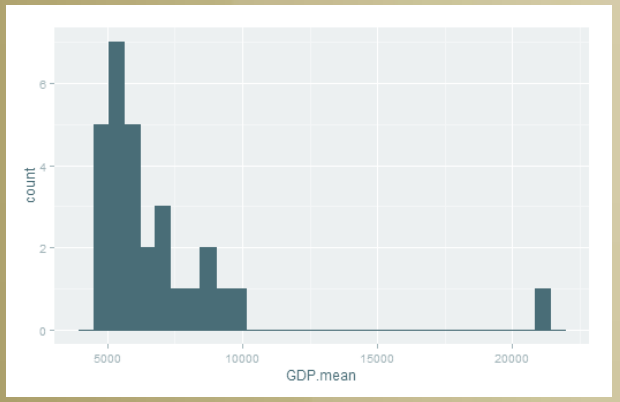
Variables Collected (3)

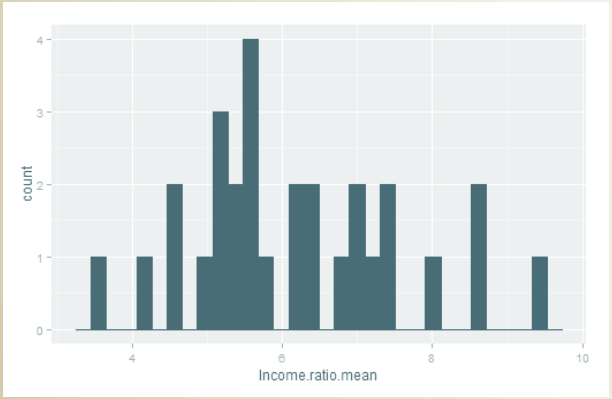
Variable	Description and Interpretation	Unit
Railway Network Density	The density of all railway lines between stations of places indicated as independent points of departure and arrival of trains carrying passengers and cargo, excluding urban railway lines. Therefore, Sofia has a low density.	Length of the road network km / 100 sq. km. of area
Share of Health Insured	The share of health insured persons as share of the population reflects the health status of the population and accessibility of health services in the district.	%
Share of Regular Internet Users	The relative share of people aged 16 to 74 that have used Internet in the past 12 months. Use of Internet also reflects access to information by the region's inhabitants, vastly improves communication and is indicative of the quality of education in the district. It should increase GDP and reduce poverty. Increased access to a great deal of information equally available also has an equalizing effect (job postings on Internet, etc.), reduces frictions and transactions costs.	%
Natural Rate of Increase	The difference between the number of annual registered live births and the annual registered number of deaths. Reflects the change of the size of the population of the region per 1000 people. Correlated with Age Dependency Ratio. Interesting to see correlations with poverty, inequality and GDP. If rich people have less children than poor people (e.g. Roma), and there are more poorer people compared to richer ones, this variable will increase inequality. On the other hand, if the poor cannot afford to have many kids, while the rich do, it will decrease inequality. The effect on GDP might also go both ways. Higher natural rates of increase will eventually increase the labor force. On the other hand, the negative natural rate of increase since he 90s have been accompanied by both periods of high GDP growth and periods with low or negative GDP growth.	Promil (1/10 of a percent)

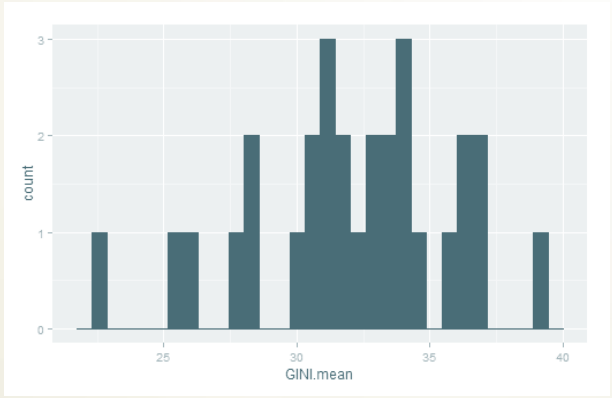
Variables Collected (4)

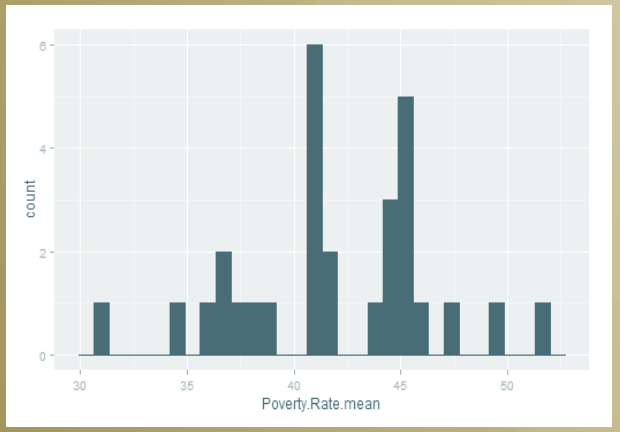
Variable	Description and Interpretation	Unit
Net Migration Rate	The difference between immigrants and emigrants to/from a region. Shows the increase or decrease of the population per 1000 people due to migration. Calculated based on the data on the number of persons who have changed their residence over the period. If poor people leave the region in search of better opportunities, while richer stay, this will decrease inequality. Also the correlation will be negative, if people tend to migrate to more equal regions consciously or not (reverse causality).	Promil (1/10 of a percent)
Age Dependency Ratio	The ratio of people aged 65+ to those aged 0-14, which are the two inactive labor market groups. A ratio too high means that for some reason the demographic structure is deteriorating. It is interesting to see how this causes or is caused by (insufficient) GDP growth, inequality and poverty levels.	%
Share Urban Population	It is interesting to see how urbanization and the concentration of population in major cities correlates with GDP level, inequality and poverty rates.	%
Share of Micro and Small Enterprises (not used)	The share of enterprises having up to, but not including 50 employees to all enterprises in the district. It is assumed that the larger share of small and medium enterprises there is in a district, the more vibrant and resistant to shocks the local economy is. Decentralization also may lead to more jobs and reduce income inequality. I used Non – Financial Companies per 1000 people instead.	%
Value Added by Factor Expenses (not used)	Indicates how much is produced in a region. A component of GDP (calculated by the production method). Due to multicolinearity and noninvertability issues when estimating, I used GDP per capita instead.	BGN

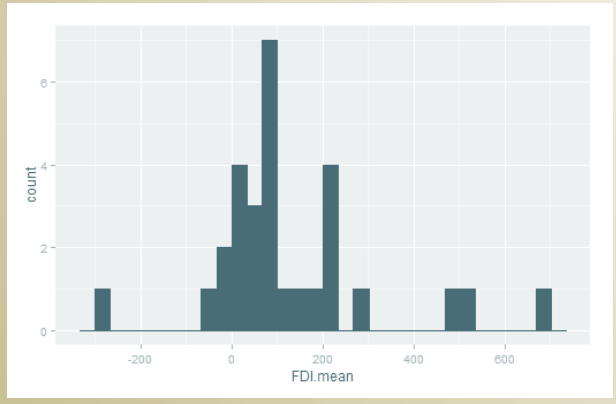
Histograms of the 5-Year (2007-2011) Averages of All Variables for All 28 Regions and Bulgaria (1)

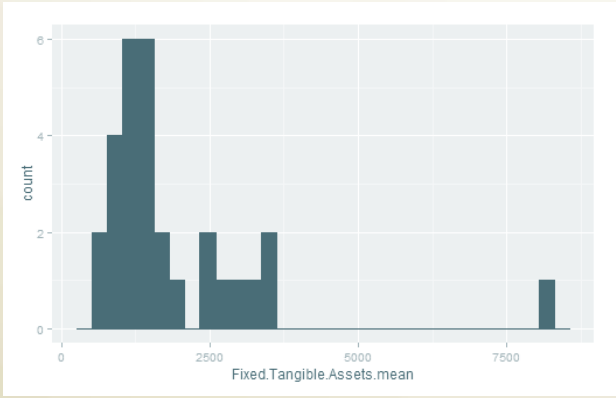


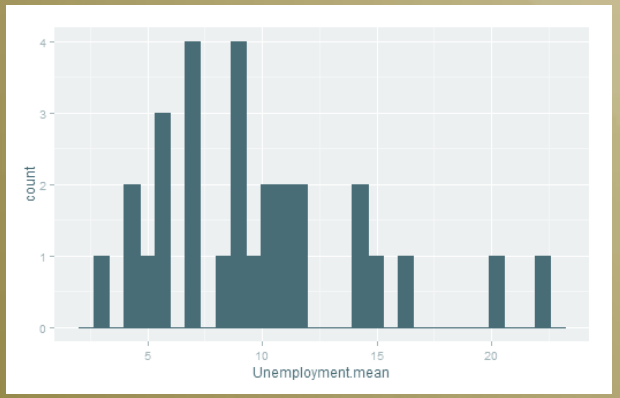


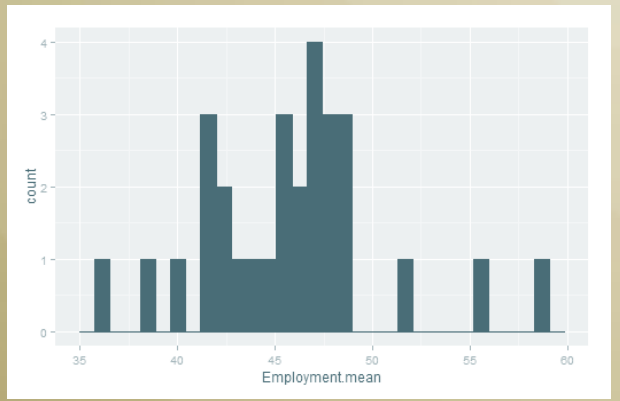


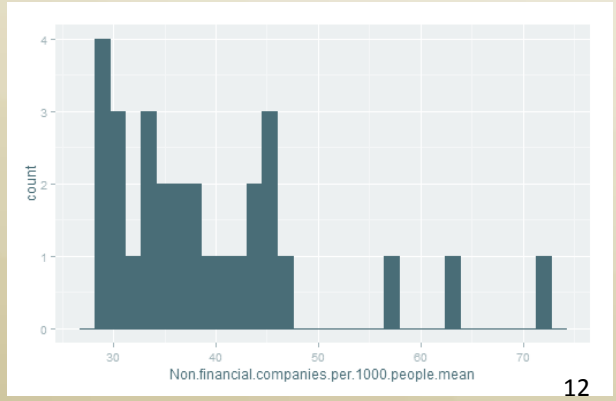




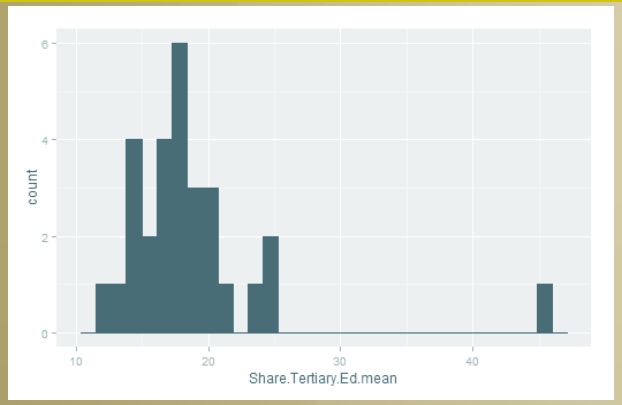


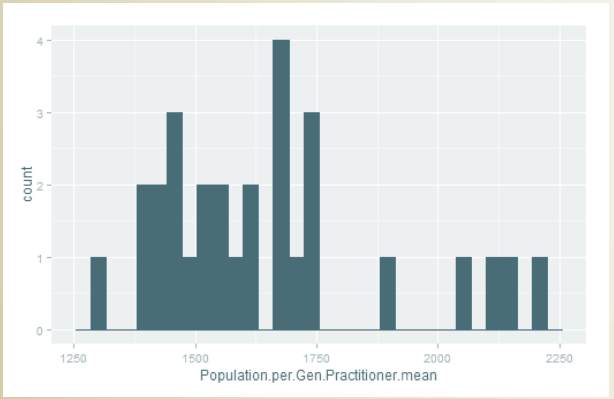


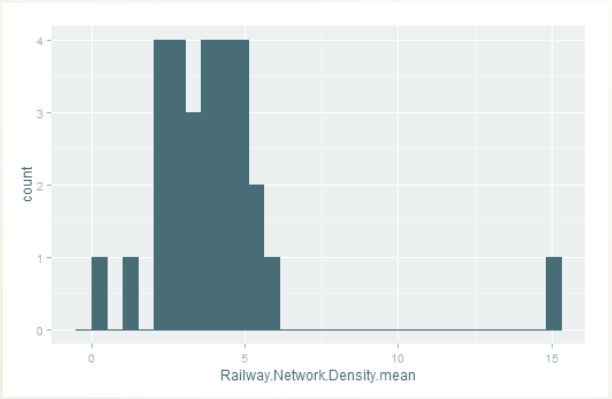


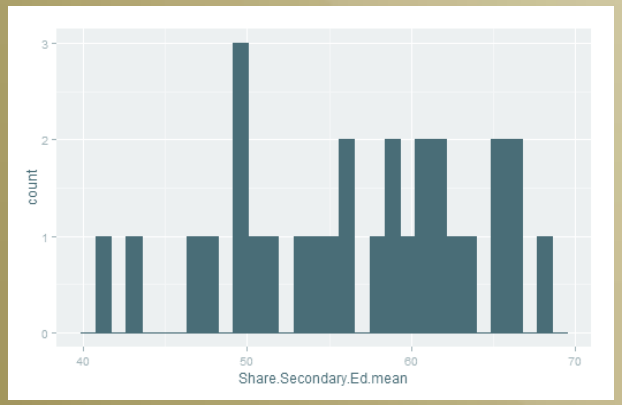


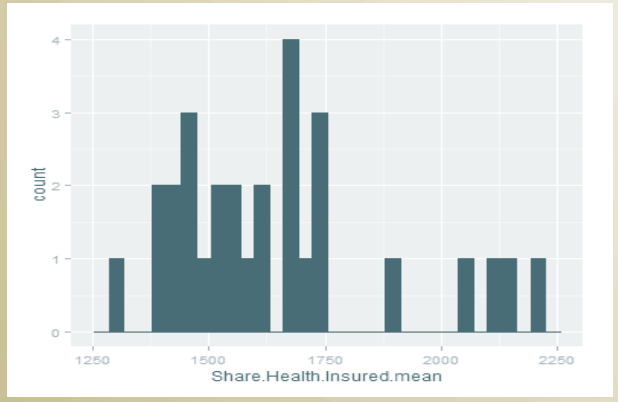
Histograms of the 5-Year (2007-2011) Averages of All Variables for All 28 Regions and Bulgaria (2)

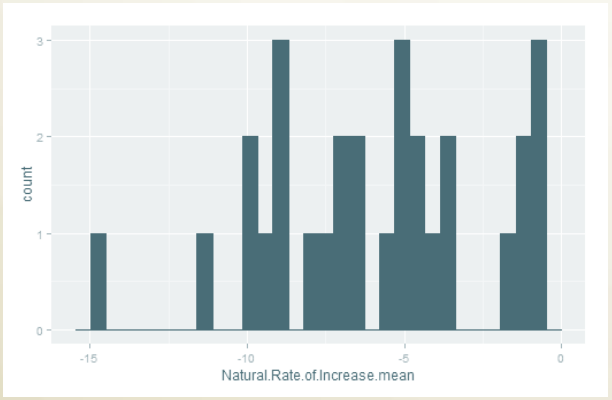


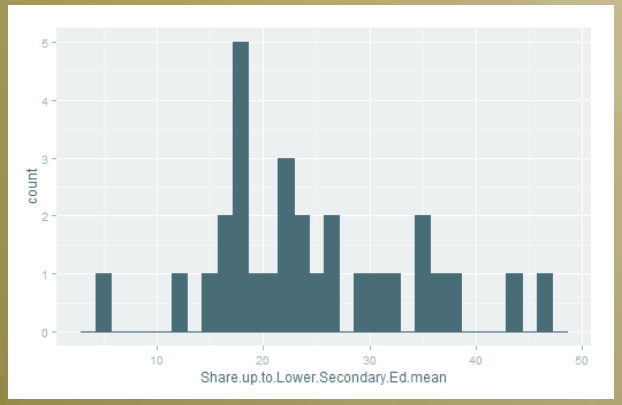


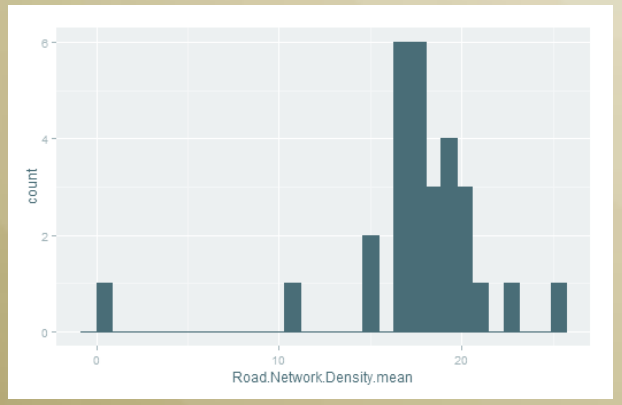


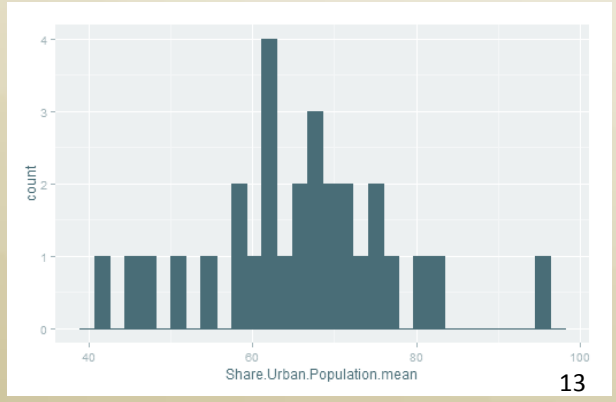




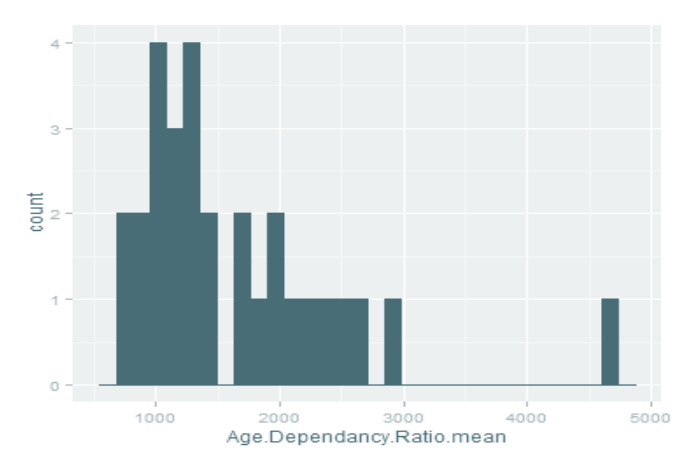


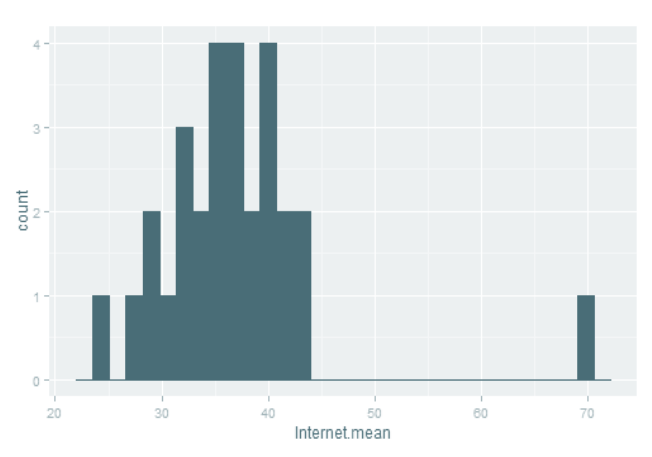


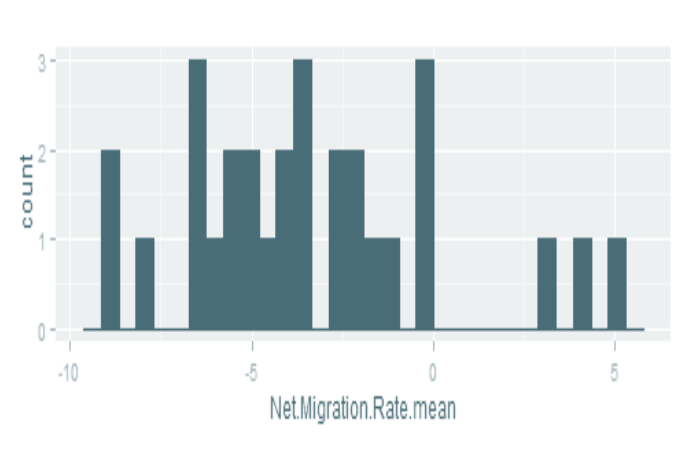


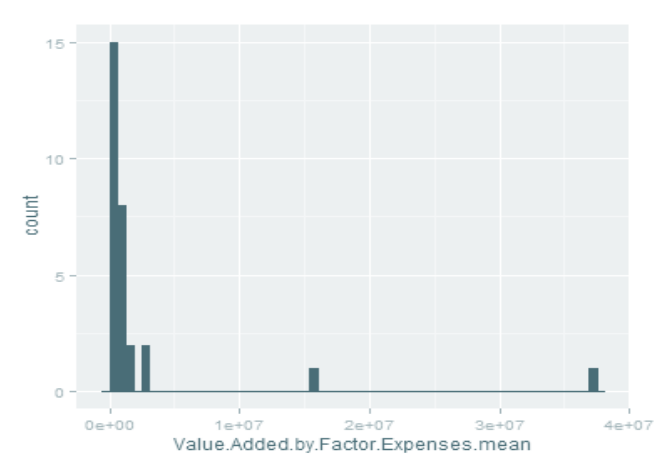


Histograms of the 5-Year (2007-2011) Averages of All Variables for All 28 Regions and Bulgaria (3)

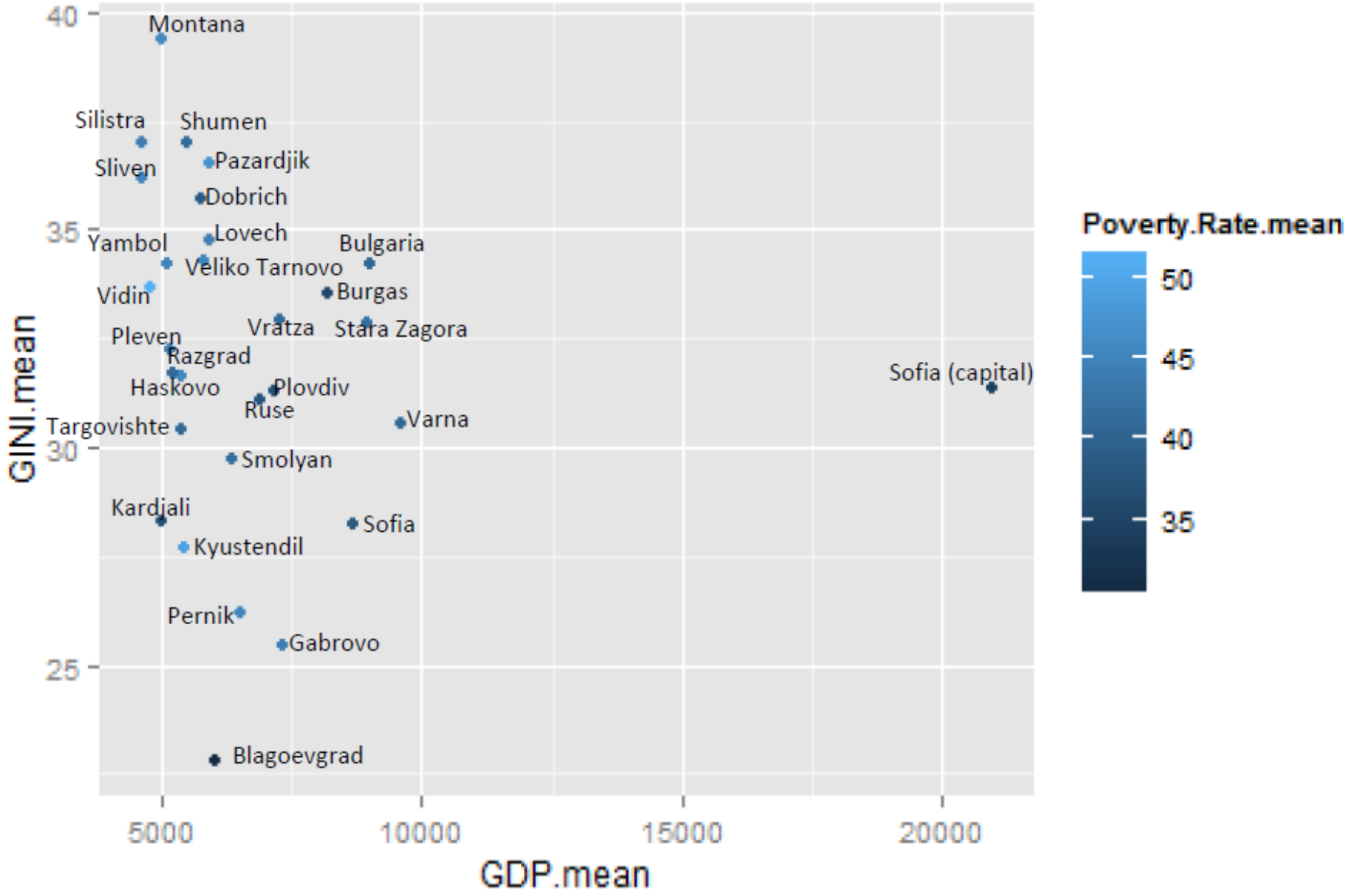






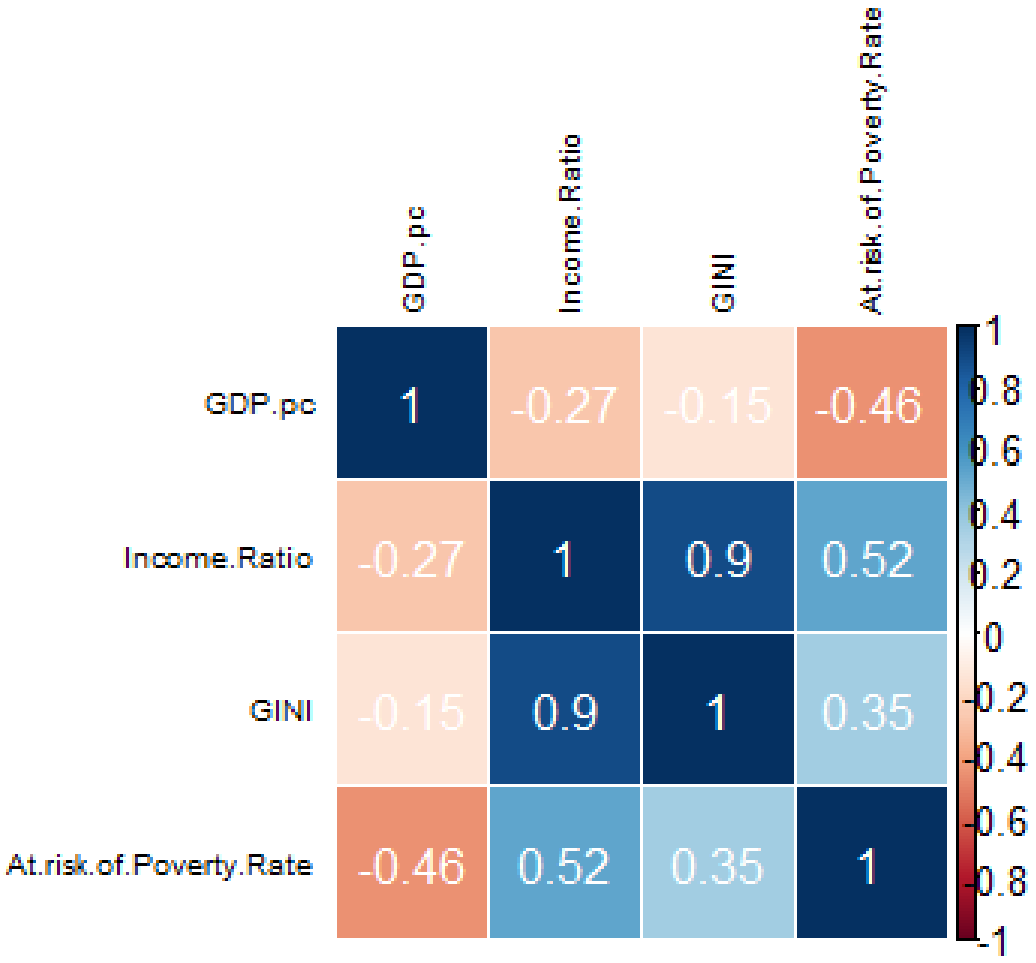


Scatterplot of the 5–Year (2007-2011) Averages of Log GDP Per Capita and GINI For All 28 Regions

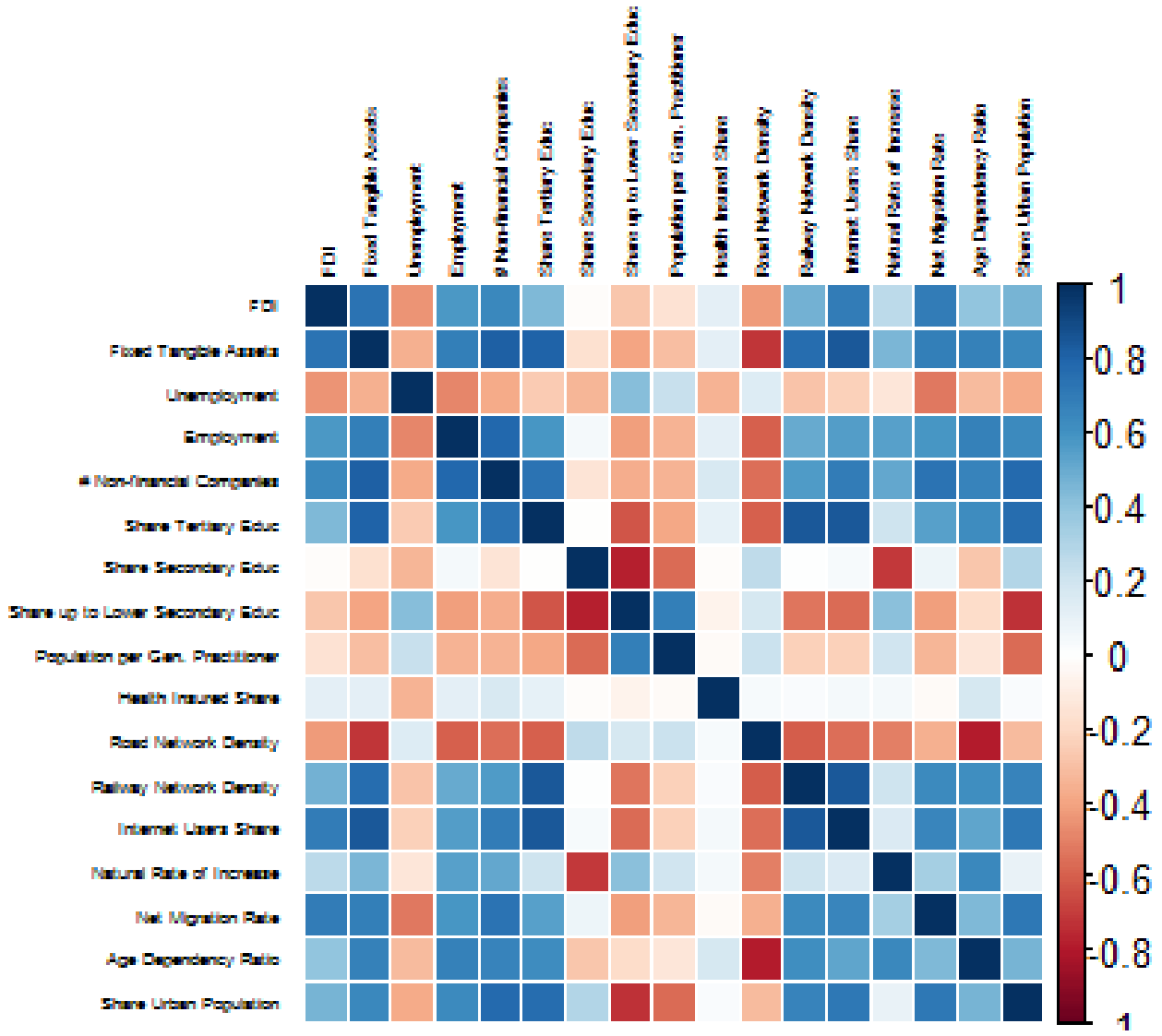


Correlations Between the 5-Year Averages of Dependent Variables Across Regions

2007-2011 Across Regions



Correlogram of All Regressors



17x17/2 = 145 correlations

97/2 correlations above abs(0.60)

29/2 correlations above abs(0.80)

Of course, this includes the 17 perfect correlations on diagonal

Choice of Variables (1)

- Regressors that could have little direct causal relationship with the dependent variables were included. However, since they could be correlated with key explanatory variables (investment, employment) and thereby have an indirect effect on the outcome variable, they need to be in the regression; else– selection bias and endogeneity problems. Conditioning on as many observable variables as possible that jointly influence a key regressor and the outcome variable removes the selection bias.
- As seen from the correlogram, multicolinearity is not a big problem. It does not make estimates inconsistent (but increases standard errors)! Besides, since there are not too many regional variables collected by NSI, overfitting the model is the lesser evil than excluding an observable variable that is correlated with a key regressor (FTA, FDI, Employment, Education).
- Value Added by Factor Expenses and Share of Micro and Small Enterprises were collected but excluded because these variables varied too little across years and across time. As a result from this, a crucial matrix in the estimator formula could not be inverted because it was singular, leading to an inability to estimate by random effects. Therefore log GDP per capita instead of the former was used in the poverty regressions and the number of non-financial enterprises per 1000 people instead of the latter (as a proxy for entrepreneurship and dynamism of the economy).
- For argumentation about why small and medium enterprises and entrepreneurship are relevant for economic growth, see for example *http://pdf.usaid.gov/pdf_docs/PNADO560.pdf*
- The unemployment rate for the inequality and poverty rate regressions and the employment rate for the GDP regressions were used.

Choice of Variables (2)

- Either natural rate of increase or age dependency ratio were used, but never both in one regression, as they have similar economic meaning and are correlated. Age dependency ratio was used in the poverty regressions because a possible causal relationship between the two has a better economic meaning in this case. On the other hand, natural rate of increase is more suited to explain GDP growth.
- Regular Use of Internet, Railway Density and Tertiary Education are all highly correlated with each other, but all of them were kept in the regressions, because they have different economic meaning. All should lead to productivity growth and greatly reduce costs, but through different channels. Railway Density and Internet both measure “interconnectedness” but through different channels.
- Different constellations of variables were tried in an effort to increase the R squared.
- The infrastructure variables in the poverty regression were not used because only non-urban roads and railways are counted, resulting in Sofia (capital), which has the lowest poverty and highest GDP, having road density of 0 and a low railway density. This gives a positive coefficient on the infrastructure variables.
- Tertiary education attainment was used in the GDP regressions (because highly educated should have the bigger role in raising GDP), secondary education attainment was used for the inequality regressions and up to lower secondary education (8th grade) for the poverty regressions. All of these choices make economic sense.

Fixed and Random Effects (1)

Rationale: Controlling for observable regressors, measured in the same period, may not be sufficient to control for confounding/endogeneity (i.e. something left in the error term correlates with one or more of the regressors, rendering the coefficients biased and inconsistent).

Solution: use time dimension of data to "difference away" or transform the problematic error component.

Two methods to do that: "fixed" and "random" effects panel regressions.

A classical question in panel economics: **Random or fixed effects?**

Unobserved effects model:

Composite error term

$$y_{it} = \beta_0 + x_{it}\beta + (c_i + u_{it}) \qquad t = 1,\ldots,T$$

- c_i unobserved component, latent variable, unobserved heterogeneity, individual effect, individual heterogeneity
- u_{it} idiosyncratic errors, idiosyncratic disturbances

Random effect: c is a random variable **un**correlated with x

Fixed effect: c is a random variable correlated with x

- c_i is an unobserved (or unmeasurable), region-specific, time-constant (hence no t index) component of the error term that causes endogeneity problems (e.g. geographical characteristics, culture…)
- u_{it} is the rest of the error term that varies both with region i and time t

Fixed and Random Effects (2)

- The fixed effects estimator uses a transformation to remove the unobserved effect c_i prior to estimation. Any time constant regressors are also removed along with it.
- The random effects estimator partially removes c_i and partially leaves it in the error term. It is used when the unobserved effect is not or weakly correlated with the regressors. This happens when we have enough good controls in our regression, and so the leftover c_i only induces serial correlation in the composite error term period to period (necessarily because errors in all times contain a time-constant c_i), but does NOT cause correlation between the composite error and the regressors. The autocorrelation does not make the coefficients inconsistent, but it does increase the standard errors and makes standard hypothesis testing incorrect. RE estimator fixes this autocorrelation problem by quasi-demeaning the error and the regressors by weighing the different observations in a different way (known as “generalized least squares”).
- If we believe that c_i is uncorrelated with the explanatory variables, the coefficients can be consistently estimated by pooled ordinary least squares (i.e. just stacking observations on top of each other, treating them as cross-sectional, ignoring the panel structure and not differencing / transforming the c_i at all, leaving it entirely in the error). But this is inefficient, we lose useful information and we have the serial correlation problem in the composite errors still (and hence invalid test statistics/standard errors). Therefore, if we assume that (1) c_i exists and (2) it is uncorrelated with the regressors (for the same region in all time periods), we use RE instead of POLS.
- Both fixed and random effects assume strict exogeneity in addition, i.e. the idiosyncratic component of the error u_{it} is uncorrelated with all regressors in all time periods. POLS does not.

Or, The Same Thing, but Formally…

POLS assumptions:

- Contemporaneous exogeneity (both parts of the error - u_{it} and c_i - are not correlated with the all regressors for the region in the **same** time period.

$$y_{it} = x_{it}\beta + v_{it} \qquad t = 1,\ldots,T$$

$$v_{it} = c_i + u_{it}$$

$$E(x_{it}'v_{it}) = 0$$

$$E(x_{it}'u_{it}) = 0$$

$$E(x_{it}'c_i) = 0$$

But POLS does not assume strict exogeneity:

$$E(u_{it}|x_{i1}, x_{i2}, \ldots, x_{iT}, c_i) = 0 \qquad t = 1, 2\ldots, T$$

$$\Rightarrow E(x_{is}'u_{it}|x_{t\neq s}, c) = 0 \qquad s, t = 1, 2\ldots, T$$

FE vs. RE assumptions

However, both FE and RE assume strict exogeneity (lack of correlations of the idiosyncratic component of the error with the regressors in **all** time periods). In addition, as mentioned, RE assumes strict exogeneity of the individual effect on top of that:

Assumption FE.1 $E(u_{it}|x_i, c_i) = 0 \quad t = 1, 2,..., T$

➤ Estimation and inference with the random effect assumption.

Assumption RE.1 (regressors not informative about mean of RE):

- $E(u_{it}|x_i, c_i) = 0 \qquad t = 1, 2,..., T \qquad x_i=(x_{i1}, \dots, x_{iT})$
- $E(c_i|x_i) = E(c_i) = 0 \qquad t = 1,..., T.$

Fixed Effects Calculation

Starting from:

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

$$y_i = X_i\beta + c_i l_T + u_i$$

where X_i is the matrix of the vectors x_{it} stacked one on another (analogously for y_i and u_i), and l_T is Tx1 vector of ones.

Taking the mean over all time periods of all variables and the error for every region and then demeaning:

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \qquad \bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}.$$

$$\underbrace{y_{it} - \bar{y}_i}_{\ddot{y}_{it}} = \underbrace{(x_{it} - \bar{x}_i)}_{\ddot{x}_{it}}\beta + \underbrace{u_{it} - \bar{u}_i}_{\ddot{u}_{it}} \qquad \text{or} \qquad \ddot{y}_{it} = \ddot{x}_{it}\,\beta + \ddot{u}_{it}$$

To estimate beta, this assumption should hold:

$$E(\ddot{x}_{it}\,'\ddot{u}_{it}) = 0 \qquad t = 1, 2, \ldots, T$$

This assumption holds under FE.1 !

Note that **time constant variables disappear** due to differencing.

FE and RE Estimator Formulas:

Assumption FE.2:

$$rank\left(\sum_{t=1}^{T} E(\ddot{x}_{it}\,'\ddot{x}_{it})\right) = rank\, E(\ddot{X}_i\,'\ddot{X}_i) = U$$

Then the FE estimator is:

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \ddot{X}_i\,'\ddot{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \ddot{X}_i\,'\ddot{y}_i\right) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}\,'\ddot{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}\,'\ddot{y}_{it}\right)$$

FE (also known as “within”) estimator is consistent under FE.1 and FE.2.

The RE estimator is:

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} X_i\,'\widehat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i\,'\widehat{\Omega}^{-1}y_i\right)$$

where $\widehat{\Omega}^{-1}$ is the estimated variance-covariance matrix of the composite error $v_{it} = c_i + u_{it}$ (how exactly it is estimated is skipped here for brevity).

Testing Model Fit (1)

- Testing for presence of unobserved effect c_i:

➤ Testing for the presence of a random effect (Breusch-Pagan test)

Null hypothesis: v_{it} are serially uncorrelated.

Test based on H_o: $\sigma_c{}^2 = 0$

Test statistic: $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it} \hat{v}_{is}$

Composite error for some region and time

Under the null, and for any distribution of v_{it}

$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it} \hat{v}_{is}$ has a limiting normal distribution with mean zero.

Testing for Model Fit (2)

- Given that we have established a presence of unobserved heterogeneity c_i, we can test whether to choose fixed or random effects.

Comparison of Estimators using the

➤ Hausman statistic:

$$\left(\hat{\delta}_{FE} - \hat{\delta}_{RE}\right)' [\mathrm{A}\,\hat{\mathrm{v}}ar\left(\hat{\delta}_{FE}\right) - \mathrm{A}\,\hat{\mathrm{v}}ar(\hat{\delta}_{RE})]^{-1}\left(\hat{\delta}_{FE} - \hat{\delta}_{RE}\right)$$

$\hat{\delta}_{RE}$ – estimated vector of RE coefficients without the coefficients on time constant variables

$\hat{\delta}_{FE}$ – estimated vector of FE coefficients (which by definition is without the coefficients on time constant variables)

This statistic is distributed as chi-squared under the RE assumptions. If it is sufficiently far from zero, i.e. the difference between the vectors of coefficients under RE and under FE is substantial, we reject the null that there is no difference and we assume that RE 1 b) assumption is false. Thus, since FE assumptions are nested within RE, we use FE. If we fail to reject the null, this is given to mean that RE 1b) is true, so it does not matter which of the two coefficients we use, but we use RE, because they are more efficient (since they use more information about the error term).

Which Regression Outputs are the Valid Ones (I Also Included the Other Two For Each Regression Because They Could Give Hints About Size and Statistical Significance of Different Variables Depending on the Assumptions of the Three Models)

	GINI Regression	Income Ratio Regression	Log GDP per capita Regression	At Risk of Poverty Rate Regression
Lagrange Multiplier Test *Null hypothesis*: • **POLS** *Alternative hypothesis*: • **Random Effects**	p-value = 0.0008271 ⇨ Reject the null ⇨ Random Effects Model is preferred	p-value = 0.001276 ⇨ Reject the null ⇨ Random Effects Model is preferred	p-value = 8.9×10^{-11} ⇨ Reject the null ⇨ Random Effects Model is preferred	p-value = 7.281×10^{-6} ⇨ Reject the null ⇨ Random Effects Model is preferred
F Test for Individual Effects *Null hypothesis*: • **POLS** *Alternative hypothesis*: • **Fixed Effects**	p-value =0.001064 ⇨ Reject the null ⇨ Fixed Effects Model is preferred	p-value = 0.00511 ⇨ Reject the null ⇨ Fixed Effects Model is preferred	p-value = 3.512×10^{-11} ⇨ Reject the null ⇨ Fixed Effects Model is preferred	p-value = 0.0001194 ⇨ Reject the null ⇨ Fixed Effects Model is preferred
Hausman Test *Null hypothesis*: • **Random Effects** *Alternative hypothesis*: • **Fixed Effects**	p-value = 0.433 ⇨ Fail to reject the null ⇨ Random Effects Model is preferred	p-value = 0.3803 ⇨ Fail to Reject the null ⇨ Random Effects Model is preferred	p-value = 9.69×10^{-10} ⇨ Reject the null ⇨ Fixed Effects Model is preferred	p-value = 0.6942 ⇨ Fail to reject the null ⇨ Random Effects Model is prefered
Conclusion: Which model is the valid one	Random Effects	Random effects	Fixed Effects	Random Effects

Table 1: Regression of the GINI Index, Using the Fixed Effects Model (aka Within Estimator)

	GINI
	Y1
	panel *linear* Within Estimation (aka Fixed Effects Model)
At Risk of Poverty.Rate	−0.007
	(0.075)
FDI in Non-Financial Enterprises per capita	0.0003
	(0.001)
Expenditures for Fixed Tangible Assets per capita	−0.001
	(0.001)
Employment Rate	0.088
	(0.195)
Non-Financial Companies per 1000 People	−0.144
	(0.093)
Share with Secondary Education	−0.240
	(0.195)
Population per General Practitioner	−0.002
	(0.005)
Health Insured Ratio	−0.218*
	(0.128)
Road Network Density	−3.558
	(4.030)
Railway Network Density	−0.030
	(2.506)
Natural Rate of Increase	−0.713
	(0.649)
Net Migration Rate	−0.059
	(0.150)
Share of Regular Internet Users	0.011
	(0.048)
Share of Urban Population	0.261
	(0.603)
Observations	135
R^2	0.134
Adjusted R^2	0.091
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression of the GINI Index, Using Random Effects Model

	GINI
	Y1
	panel *linear* Random Effects Model
At Risk of Poverty.Rate	0.041
	(0.067)
FDI in Non-Financial Enterprises per capita	0.0003
	(0.001)
Expenditures for Fixed Tangible Assets per capita	−0.0003
	(0.0005)
Employment Rate	−0.181
	(0.128)
Non-Financial Companies per 1000 People	−0.160**
	(0.065)
Share with Secondary Education	−0.260**
	(0.108)
Population per General Practitioner	−0.003
	(0.003)
Health Insured Ratio	−0.364***
	(0.102)
Road Network Density	−0.340*
	(0.190)
Railway Network Density	−0.321
	(0.360)
Natural Rate of Increase	−0.139
	(0.270)
Net Migration Rate	−0.052
	(0.119)
Share of Regular Internet Users	0.034
	(0.041)
Share of Urban Population	0.176*
	(0.092)
Constant	89.970***
	(13.260)
Observations	135
R^2	0.366
Adjusted R^2	0.326
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: Regression of the GINI Index, Using Pooled Ordinary Least Squares

	GINI
	Y1
	panel
	linear
	Pooled Ordinary Least Squares
At Risk of Poverty.Rate	0.089
	(0.067)
FDI in Non-Financial Enterprises per capita	0.0004
	(0.001)
Expenditures for Fixed Tangible Assets per capita	0.0002
	(0.0004)
Employment Rate	-0.320^{***}
	(0.116)
Non-Financial Companies per 1000 People	-0.167^{***}
	(0.059)
Share with Secondary Education	-0.223^{**}
	(0.088)
Population per General Practitioner	-0.003
	(0.002)
Health Insured Ratio	-0.459^{***}
	(0.091)
Road Network Density	-0.298^{**}
	(0.142)
Railway Network Density	-0.414
	(0.268)
Natural Rate of Increase	0.037
	(0.209)
Net Migration Rate	-0.082
	(0.114)
Share of Regular Internet Users	0.055
	(0.043)
Share of Urban Population	0.177^{**}
	(0.072)
Constant	100.236^{***}
	(11.210)
Observations	135
R^2	0.425
Adjusted R^2	0.378
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Regression of the Income Ratio, Using the Fixed Effects Model (aka Within Estimator)

	Income Ratio
	Y2
	panel *linear* Within Estimation (aka Fixed Effects Model)
At Risk of Poverty.Rate	0.021
	(0.029)
FDI in Non-Financial Enterprises per capita	0.0002
	(0.0003)
Expenditures for Fixed Tangible Assets per capita	−0.0004
	(0.0002)
Employment Rate	0.078
	(0.074)
Non-Financial Companies per 1000 People	−0.035
	(0.036)
Share with Secondary Education	−0.093
	(0.074)
Population per General Practitioner	−0.001
	(0.002)
Health Insured Ratio	−0.044
	(0.049)
Road Network Density	−1.796
	(1.537)
Railway Network Density	−0.488
	(0.956)
Natural Rate of Increase	−0.417*
	(0.248)
Net Migration Rate	0.009
	(0.057)
Share of Regular Internet Users	0.018
	(0.018)
Share of Urban Population	0.012
	(0.230)
Observations	135
R^2	0.119
Adjusted R^2	0.081
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Regression of The Income Ratio, Using Random Effects Model

	Income Ratio
	Y2
	panel *linear* Random Effects Model
At Risk of Poverty.Rate	0.062**
	(0.025)
FDI in Non-Financial Enterprises per capita	0.0002
	(0.0003)
Expenditures for Fixed Tangible Assets per capita	−0.0001
	(0.0002)
Employment Rate	−0.088*
	(0.046)
Non-Financial Companies per 1000 People	−0.074***
	(0.024)
Share with Secondary Education	−0.050
	(0.037)
Population per General Practitioner	−0.001
	(0.001)
Health Insured Ratio	−0.102***
	(0.037)
Road Network Density	−0.159**
	(0.063)
Railway Network Density	−0.241**
	(0.118)
Natural Rate of Increase	0.019
	(0.090)
Net Migration Rate	−0.005
	(0.044)
Share of Regular Internet Users	0.025
	(0.016)
Share of Urban Population	0.071**
	(0.031)
Constant	22.127***
	(4.660)
Observations	135
R^2	0.336
Adjusted R^2	0.299
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Regression of the Income Ratio, Using Pooled Ordinary Least Squares

	Income Ratio
	Y2
	panel
	linear
	Pooled Ordinary Least Squares
At Risk of Poverty.Rate	0.078^{***}
	(0.025)
FDI in Non-Financial Enterprises per capita	0.0001
	(0.0003)
Expenditures for Fixed Tangible Assets per capita	0.0001
	(0.0002)
Employment Rate	-0.122^{***}
	(0.043)
Non-Financial Companies per 1000 People	-0.078^{***}
	(0.022)
Share with Secondary Education	-0.036
	(0.033)
Population per General Practitioner	-0.001
	(0.001)
Health Insured Ratio	-0.117^{***}
	(0.034)
Road Network Density	-0.156^{***}
	(0.053)
Railway Network Density	-0.272^{***}
	(0.099)
Natural Rate of Increase	0.067
	(0.077)
Net Migration Rate	-0.013
	(0.042)
Share of Regular Internet Users	0.029^{*}
	(0.016)
Share of Urban Population	0.074^{***}
	(0.027)
Constant	23.342^{***}
	(4.157)
Observations	135
R^2	0.409
Adjusted R^2	0.364
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Regression of the log GDP per capita Using the Fixed Effects Model (aka Within Estimator)

	log GDP per capita
	Y4
	panel *linear* Within Estimation (aka Fixed Effects Model)
FDI in Non-Financial Enterprises per capita	0.0001
	(0.0002)
Expenditures for Fixed Tangible Assets per capita	0.0001***
	(0.00004)
Employment Rate	0.015*
	(0.008)
Non-Financial Companies per 1000 People	−0.002
	(0.008)
Share with Tertiary Education	0.001
	(0.012)
Health Insured Ratio	0.012
	(0.009)
Road Network Density	0.018**
	(0.008)
Railway Network Density	0.003
	(0.031)
Natural Rate of Increase	−0.001
	(0.009)
Net Migration Rate	0.004
	(0.014)
Share of Regular Internet Users	0.014
	(0.008)
Share of Urban Population	−0.004
	(0.005)
Observations	29
R^2	0.958
Adjusted R^2	0.363
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Regression of the log GDP per capita Using the Random Effects Model

	log GDP per capita
	Y4
	panel *linear* Random Effects Model
FDI in Non-Financial Enterprises per capita	0.00004
	(0.0002)
Expenditures for Fixed Tangible Assets per capita	0.0002***
	(0.00003)
Employment Rate	0.019**
	(0.007)
Non-Financial Companies per 1000 People	−0.002
	(0.005)
Share with Tertiary Education	0.002
	(0.009)
Health Insured Ratio	0.011*
	(0.006)
Road Network Density	0.016**
	(0.007)
Railway Network Density	−0.005
	(0.020)
Natural Rate of Increase	−0.007
	(0.008)
Net Migration Rate	0.016
	(0.010)
Share of Regular Internet Users	0.010
	(0.006)
Share of Urban Population	−0.005
	(0.004)
Constant	6.370***
	(0.572)
Observations	29
R^2	0.960
Adjusted R^2	0.529
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Regression of the log GDP per capita Using Pooled Ordinary Least Squares

	Dependent variable:
	Y4
	panel linear
FDI in Non-Financial Enterprises per capita	0.00004
	(0.0002)
Expenditures for Fixed Tangible Assets per capita	0.0002***
	(0.00003)
Employment Rate	0.019**
	(0.007)
Non-Financial Companies per 1000 People	−0.002
	(0.005)
Share with Tertiary Education	0.002
	(0.009)
Health Insured Ratio	0.011*
	(0.006)
Road Network Density	0.016**
	(0.007)
Railway Network Density	−0.005
	(0.020)
Natural Rate of Increase	−0.007
	(0.008)
Net Migration Rate	0.016
	(0.010)
Share of Regular Internet Users	0.010
	(0.006)
Share of Urban Population	−0.005
	(0.004)
Constant	6.370***
	(0.572)
Observations	29
R^2	0.960
Adjusted R^2	0.529
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Regression of the At Risk of Poverty Rate, Using Fixed Effects Model (aka Within Estimator)

	People at Risk of Poverty Rate
	Y3
	panel
	linear
	Fixed Effects Model (aka Within Estimation)
GDP per capita	−0.00002
	(0.001)
FDI in Non-Financial Enterprises per capita	0.0002
	(0.001)
Expenditures for Fixed Tangible Assets per capita	−0.001
	(0.001)
Unemployment Rate	0.043
	(0.191)
Non-Financial Companies per 1000 People	0.073
	(0.131)
Share with Secondary Education	−0.019
	(0.244)
Population per General Practitioner	−0.007
	(0.008)
Share.Health.Insured	−0.019
	(0.183)
Age Dependancy Ratio	−0.018
	(0.017)
Net Migration Rate	0.004
	(0.213)
Internet	−0.007
	(0.068)
Share of Urban Population	−0.978
	(0.801)
Observations	135
R^2	0.094
Adjusted R^2	0.065
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: Regression of the At Risk of Poverty Rate, Using Random Effects Model

	People at Risk of Poverty Rate
	Y3
	panel *linear* Random Effects Model
GDP per capita	0.0001
	(0.0004)
FDI in Non-Financial Enterprises per capita	−0.0001
	(0.001)
Expenditures for Fixed Tangible Assets per capita	−0.001*
	(0.001)
Unemployment Rate	0.199*
	(0.118)
Non-Financial Companies per 1000 People	−0.098
	(0.093)
Share with Secondary Education	0.031
	(0.125)
Population per General Practitioner	−0.004
	(0.004)
Share.Health.Insured	0.033
	(0.142)
Age Dependancy Ratio	−0.001
	(0.001)
Net Migration Rate	−0.119
	(0.166)
Internet	0.003
	(0.058)
Share of Urban Population	0.070
	(0.113)
Constant	43.865**
	(18.520)
Observations	135
R^2	0.265
Adjusted R^2	0.239
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 12: Regression of the At Risk of Poverty Rate, Using Pooled Ordinary Least Squares

	People at Risk of Poverty Rate
	Y3
	panel
	linear
	Pooled Ordinary Least Squares
GDP per capita	0.0004
	(0.0004)
FDI in Non-Financial Enterprises per capita	−0.0004
	(0.001)
Expenditures for Fixed Tangible Assets per capita	−0.001**
	(0.001)
Unemployment Rate	0.244**
	(0.107)
Non-Financial Companies per 1000 People	−0.198**
	(0.077)
Share with Secondary Education	0.045
	(0.100)
Population per General Practitioner	−0.003
	(0.003)
Share.Health.Insured	−0.003
	(0.128)
Age Dependancy Ratio	−0.0004
	(0.001)
Net Migration Rate	−0.237
	(0.158)
Internet	0.015
	(0.064)
Share of Urban Population	0.153*
	(0.078)
Constant	39.341**
	(15.613)
Observations	135
R^2	0.343
Adjusted R^2	0.310
Note:	*p<0.1; **p<0.05; ***p<0.01

Summary of Results

Regression	Model Preferred	Statistically Significant Variables	Size of the effect
GINI Regression	Table 2 (Random)	• **Number of Non-Financial Companies per 1000 people** • Share With Secondary Education • **Health Insured Ratio** • **Road Network Density** • **Share of Urban Population**	• -0.16 (decrease ineq.) • -0.26 • -0.364 • -0.340 • 0.176
Income Ratio Regression	Table 5 (Random)	• At Risk of Poverty Rate • Employment rate • **Number of Non-Financial Companies per 1000 people** • **Health Insured Ratio** • Road Network Density • Railway Network Density • **Share of Urban Population**	• 0.062 • -0.088 • -0.074 • -0.102 • -0.159 • -0.241 • 0.071
Log GDP per capita Regression	Table 7 (Fixed)	• Expenditures for Fixed Tangible Assets per Capita • Employment Rate • Road Network Density • (only in Random Effects) Health Insured Ratio The same variables are statistically significant in the POLS regression, with similar coefficients	• 0.0001(negligible) • 0.015 • 0.018 • 0.011
At Risk of Poverty Rate Regression	Table 11 (Random) In Table 12 (POLS), in addition to the first 2 variables	• Expenditures for Fixed Tangible Assets per Capita • Unemployment rate • Non-Financial Companies per 1000 People • Share of Urban Population	• -0.001 (negligible) • 0.199 • -0.198 • 0.153

Policy Conclusions (1)

- FDI in non-financial enterprises and expenditures for fixed tangible assets are **both** statistically insignificant and have **extremely small coefficients** on top of that in all six inequality regressions.
- But employment has a larger, negative and statistically significant effect on inequality. Therefore, to reduce inequality, contrary to popular wisdom, regions do not need just any foreign investments or tangible assets => they need to be job-creating!
- Shopping centers, malls, photovoltaics…
- Bulgaria Invest Agency should implement policies encouraging job-creating (foreign) investments in times of capital inflow and economic boom.
- Think about what the word "investment" should mean.
- Health Insured Ratio is statistically significant, largely reduces inequality and raises GDP. Regions with more equal income tend to be more health insured. If people on an equal playing field, they are more prone to contribute to such schemes.
- Instead of investment, number of non-financial enterprises per 1000 people is statistically significant and has a LARGE effect in reducing inequality (1 more firm decreases GINI by 0.16)
- => Fostering of entrepreneurship, development of SMEs, especially job-creating ones.
- Access to credit, low interest rates, Development Bank, Insurance and Risk Management Schemes, JEREMIE, Business incubators (but not only in IT and not only in Sofia).
- Much greater control on corruption on a regional level, faster and unbiased judiciaries on a local level, fewer regulations and permits to open and operate businesses; equality in front of the law, level playing field

Policy Conclusions (2)

- Focus on less capital intensive (that is not requiring huge sunk investments) and *productive* industries with big export and employment potential (due to the importance of employment in raising GDP and reducing inequality and poverty). In addition, such industries are the ones in which Bulgaria has competitive advantages in: agriculture, healthy foods, (cultural) tourism, winemaking, IT, arts. The direction in the last several years is the right one! Also these are industries that are less prone to the economic cycle and speculative credit bubbles.
- When deciding on funding specific SMEs and projects, not everything that sounds "innovative" and "trendy" is actually job-creating. A change of paradigm with less focus on "innovation" and more on "common sense", marketplace needs and jobs. Balanced focus on 7-8 industries, not just one or two (e.g IT) to avoid herding effect in business plans applying and in financing.
- Policies to encourage the private banking sector to give credits to SMEs in these priority areas (tax breaks, subsidies, etc.), instead of relying only on state development banks.
- However, the author strongly believes the government's role is only to make certain investments more attractive or unattractive; the investments should be from the private sector.

Policy Conclusions (3)

- The same variables show up statistically significant in both the **inequality, GDP** and **poverty rate** regressions: (un)employment, entrepreneurship and SMEs, (road) infrastructure. By encouraging job-creative SMEs, we solve all three of the above problems.
- Infrastructure development is crucial: right priorities in the last several years!
- In the three GDP regressions, again, the investment in FTA p.c. and FDI in non-financial enterprises p.c. are both insignificant statistically and have extremely small sizes. However, employment is statistically significant in all tables 7-9. Therefore FDI and FTA are important inasmuch as they provide employment. This is a quite unexpected result for GDP (expected for inequality and poverty rate)! Probably due to the regional focus of the study.
- In the GDP regressions, the entrepreneurship variable has a negative effect; however, it is negligibly small and is not statistically significant, so no problem with this.
- Share of population with secondary education reduces HUGELY GINI (but not the income ratio) => policies to increase the number of people graduating high school.
- Urbanization increases inequality and is statistically significant. If job-creating SMEs are increasingly funded in less-developed areas, (which is the main policy which should solve most of the problems), urbanization will decrease as a side effect and will further decrease inequality.
- Tertiary education's effect on GDP is very low and statistically insignificant. Again, only important inasmuch as it leads to employment. Implication: education reform, surveys for the businesses to gather data on what specialties is needed, co-op/part-work/study programs, practical education.

Policy Conclusions (4)

- Surprisingly to the author, demographics do not have large or statistically significant effects on either GDP or the at risk of poverty rate. In the GDP regressions, the natural rate of increase has a negative effect on GDP per capita. Since GDP is measured in per capita terms, reduction of population (negative NRI) increases GDP per capita arithmetically (as it decreases the denominator in the GDP per capita calculation) if economically inactive people pass away, hence the negative correlation between the two. This is counteracted, however, by the possible reduction of GDP due to the reduced population and consequently labor force (assuming that the other two factors of production – total factor productivity and capital - stay the same or do not increase enough to compensate the reduction of labor), so the overall size of the coefficient is small and the result is inconclusive. However, effects of demographics are very slow to give effect on **current** GDP (which the regression uses), so these inconclusive results are understandable. However, they are sure to have negative effect in the medium term future. Furthermore, the years of high GDP growth and low unemployment in Bulgaria were characterized by very low natural rates of increase. Roma people have a lot of children, whose number is not correlated with the economic cycle. If many of them are economically inactive and do not contribute substantially to GDP, their large natural rate of increase will have decreasing effect on GDP per capita, so this again explains the sign. Hence, well-thought out policies for informed parenthood, integration and inclusion etc. are called for.
- Net migration rate is not statistically significant in any of the 12 regressions. However, if we believe the signs in the insignificant results, people migrate to a greater extent to more equal regions with higher GDP and less poverty, which makes sense.
- The single most important variable explaining the risk of poverty rate is unemployment. The others are important inasmuch as they influence unemployment => job-creating policies again.

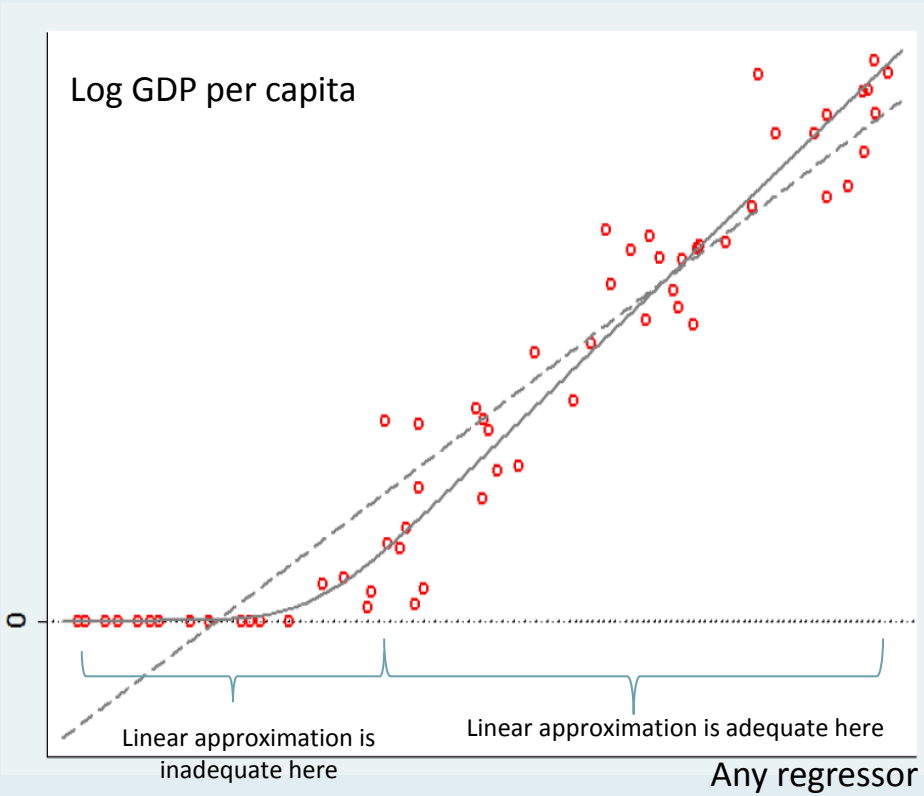
Econometric Conclusions

- Comparatively low R-squareds for the inequality regressions that are valid (Table 2 and Table 5). Therefore, inequality is explained by some variables still that are not measured. It would be great if they were available to policymakers to make informed decisions.
- Possible suggestions: Utilization of EU funds by sectors and by regions, jobs created for each project receiving funding to gauge whether these funds really trickle down to the population as a whole.
- Very hard to measure the really relevant variables for inequality: bribery, unofficial payments, corruption of local administration and local judicial systems. Inequality us as much an economic as an institutional problem.
- There is a significant jump in R-squareds going from fixed effects to random effects to pooled OLS.
- There is a tendency for the absolute value of the coefficient on the same variable to decline from POLS to Random Effects to Fixed Effects. This is because POLS leaves the unobserved heterogeneity in the error term, hence the coefficients left in the regression incorrectly incorporate its effect too. Fixed effect completely differences it away, so no interference from it is possible.
- The econometric assumptions matter a lot because not only some variables can become statistically significant, but also the signs can sometimes change between models! (e.g. employment sign in regressions 1-3)
- But in general there is not a big difference between the significance of the coefficients (a lot of variables that are statistically significant in the FE, remain so in RE, but more rarely vice versa).
- The R squared for all the GDP regression is 0.96 ! This means that the GDP model with the regressors available captures a huge amount of variability of GDP.
- The available variables are inadequate in explaining the poverty rate. Low R-squareds, some strange signs (which are not statistically significant). Therefore, the model needs to be refined.

Ideas for Future Research

- EU funds utilization's role on poverty reduction, inequality. Which programs and projects have reduced it the greatest? Which affect only GDP, but not poverty or inequality?
- Collecting statistical data for expenditures for Fixed Tangible Assets by economic sectors and regions when they become available: now available only for 2010-2012 (the author was not able to examine their impact on inequality and poverty rate due to little overlap with time series for poverty and inequality 2007-2011). Rationale: investments in which fields should be fostered? Investments in which sectors have decreased inequality and increased employment the most?
- Improvement of the models: Nonlinear regressions and looking for the correct functional form (including quadratics, interactions, running probit and tobit regressions). Inequality, poverty rate and GDP levels cannot go negative, so linear forms are incorrect for values close to 0 (slope of the regression line should gradually go to zero there). However, values close to GDP=0, risk of poverty = 0 and GINI=0 are unrealistic. On the other hand, the flattening of the slopes could start far from these points and render the linear approximation inadequate. At risk of poverty levels and income ratio are also bounded from above (at 1), therefore the regression line should be S-shaped. A parabola example of log GDP (bounded by 0 only) is shown below.

Non-Linear vs. Linear



Bozhidar Karanovsky , University of St.Gallen, MiQEF

E-mail: bozhidar.karanovsky@gmail.com
bozhidar.karanovsky@student.unisg.ch

Thank you for your time!