SKY IS THE LIMIT?

Measuring economic growth and regional income inequality with nighttime lights data and machine learning

Tufts UNIVERSITY
M. Ozgur BOZCAGA
UEP 232, December 2019

MOTIVATION

The early stage of high economic growth is usually coupled with high inequality. This study examines the trends in regional inequality in a developing country, Turkey, which experienced rapid growth in the last three decades using a novel estimation method combining nighttime lights (NTL) data, spatial analysis, and machine learning.

While national accounts are the primary source of data for any research on growth and inequality, usually developing countries lack data at the granularity required for sub-regional analysis. In this regard, using NTL as a proxy for regional economic activity, poverty, and national income has long been practiced in the literature.

However, without reliable benchmark data at the local level, it is not possible to test the validity of the income and economic activity estimates that rest on NTL.

Also, using a single value (such as mean or maximum luminosity) from NTL data for a single administrative unit might lead to biased estimates. To overcome these issues, I use LASSO to determine the weights that should be assigned to each NTL summary statistics value when predicting level of development and income at the local level. Then, I estimate development and income values for districts (sub-provincial administrative units) in Turkey between 1992-2012 and trends in overall and inter-regional income inequality.

RESEARCH QUESTIONS

1) Can we improve the degree of alignment between income and development measured by national accounts and estimated by NTL using machine learning methods?

2) Where are the clusters of high and low economic growth located in Turkey between 1992-2012?

3) How did overall and regional income inequality evolved over time in Turkey between 1992-2012?

METHODS

Using annual NTL data from DMSP-OLS stable lights time-series might lead to biased results if we include light sources from non-residential or non-industrial areas for our estimates of local income or development. To avoid this potential bias, I used CORINE land cover (CLC) data to implement raster clip to extract urban and rural residential areas, industrial zones, airports, and seaports from the DMSP-OLS data for years 1992, 2000 and 2012. I also validated the extent of residential zones in CLC data with a higher resolution urban footprint data from Global Urban Footprint (GUF).

With zonal statistics tool, I calculated summary statistics (mean, median, minimum, maximum, and sum) for NTL pixel values and additionally calculated average luminosity in each district and province overtime.

To generate estimates for income in each district, I used Turkey's 2004 District Development Index as a benchmark and output from zonal statistics and average luminosity as predictors for the development index values. Instead of using an ordinary least squares estimation, I used a regression analysis method in machine learning called LASSO for selecting variables and estimating parameters to predict the development index values.

Using LASSO as the estimation method avoids overfitting and adding irrelevant variables to the model, which improved the out-of-sample performance of the prediction. The estimated model suggested that the mean and maximum luminosity variables are sufficient to predict development index values.

Using parameters for mean and maximum luminosity, I predicted development index values for 1992, 2000, and 2012, and used those values as proxies for income at the district level.

To measure income inequality, I used the two-stage Theil decomposition method, which allows the calculation of an index of overall inequality and the contribution of each component to this overall measure. The components are inequality within provinces, inequality between provinces and inequality between regions.

For this analysis, I took seven geographical regions of Turkey as regions where provinces are nested in, and the basic regional units are districts. I use the population of regions, provinces, and districts in Turkey in 1990, 2000 and 2012, and the sum of all pixel values in a given administrative unit as the NTL variable. This method weights the sum of all pixel values by the population of each administrative unit.

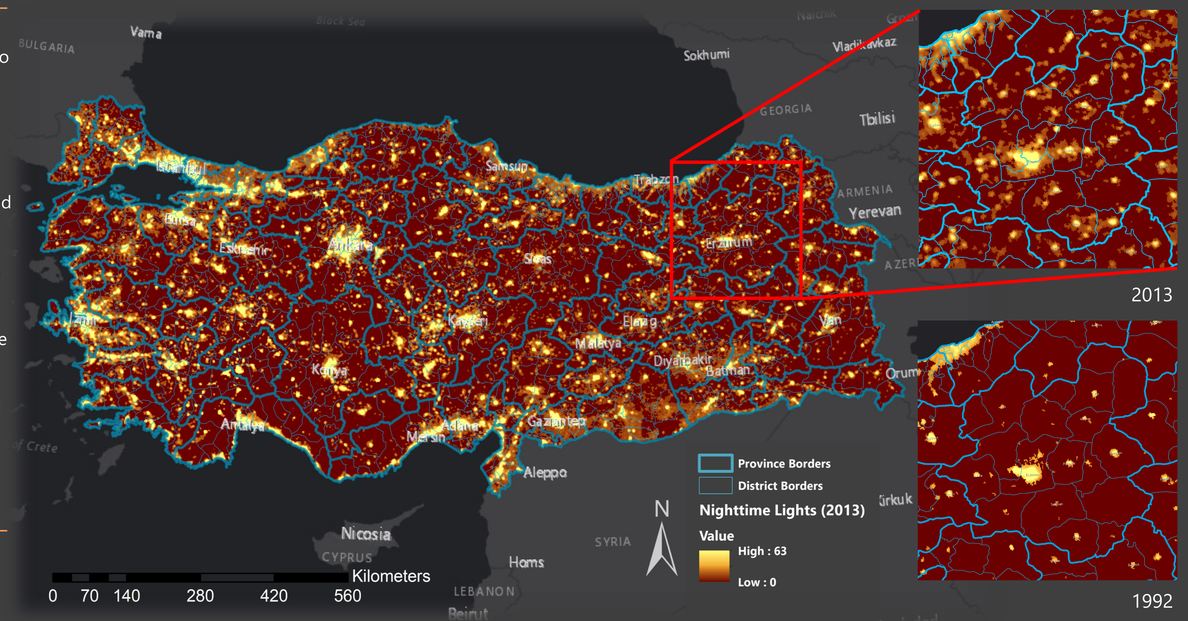

Figure 1: Nighttime lights luminosity in Turkey's provinces and districts in 2013 from DMSP-OLS NTL Time Series, Average Stable Lights (left) Nighttime lights luminosity in the districts of Erzurum province and surrounding districts in 1992 (top-right) and 2013 (bottom-right)

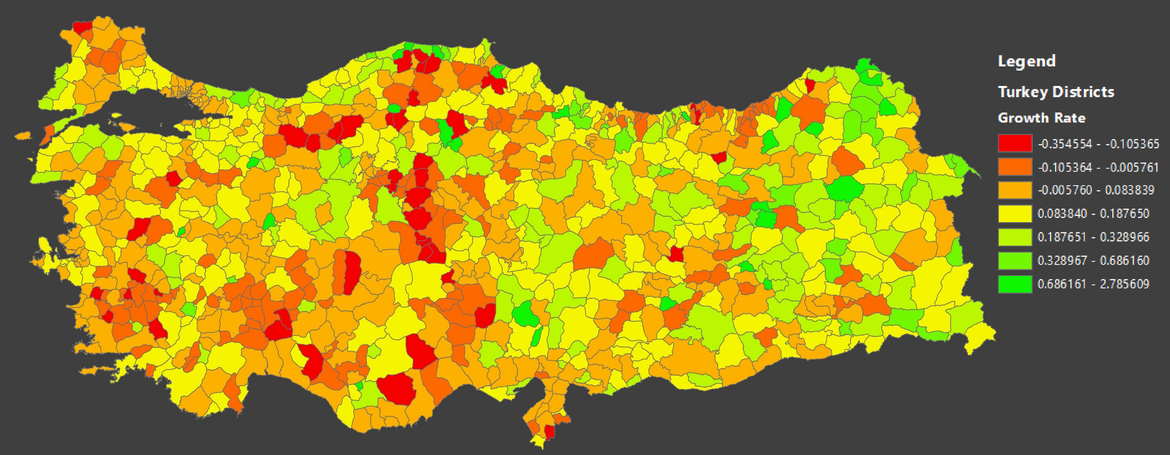

Figure 2: Average predicted real income growth in Turkey's districts between 1992—2012

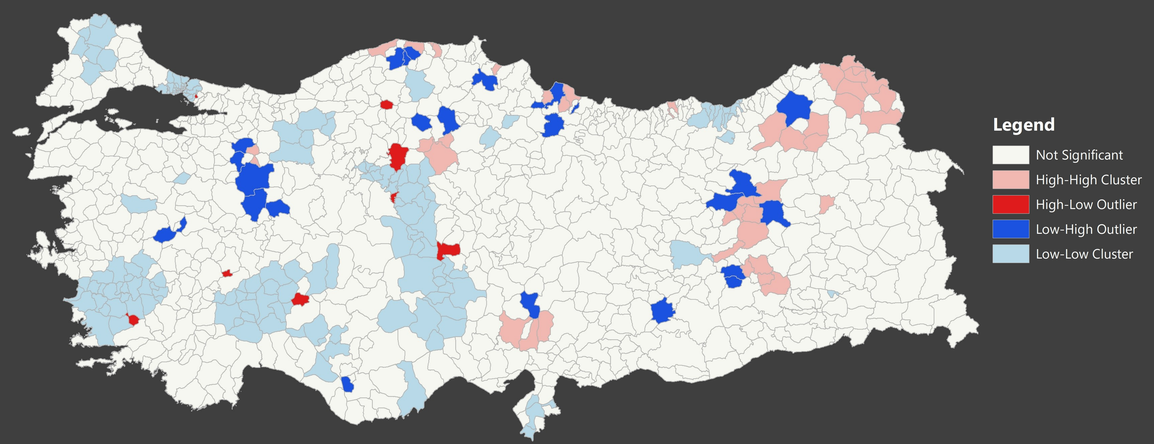

Figure 3: Clusters of economic growth in Turkey's districts between 1992—2012

FINDINGS

Figure 1 focuses on a particular province in eastern Turkey, Erzurum, which had one of the starkest increases in NTL between 1992 and 2013 when compared with other provinces. This change is also reflected in the overall economic growth in the eastern part of the country.

In Figure 2, I show the change in real (inflation adjusted) income growth rate. First, the eastern and the south-eastern regions of Turkey enjoyed above-average economic growth between 1992 -2012. This might be an outcome of the decrease in separatist armed conflict in the region and intensive infrastructure investment by the central government Second, a striking finding is the decline in economic activity in central Anatolia which had housed substantial industry investments from state-owned enterprises due to its strategic location in the latter half of the 20th century (hard to invade during Cold War) and proximity to the capital, Ankara. As the presence of the state in industrial production waned after 1990s these areas lost its importance, and residents started to emigrate to larger cities for better employment opportunities.

Third, major cities like Istanbul, Ankara and Izmir experienced growth rates that are in line with inflation rates and real growth around zero percent. This might be due to biases in estimation due to already high mean and max values in 1992.

Local Moran's I analysis results in Figure 3 validates the overall findings from Figure 2. Clusters in Eastern Turkey stand out as areas of high economic growth, whereas Central Anatolia is the most noticeable low economic growth cluster in the country.

Figure 4 graphically presents the results of the temporal inequality analysis. It shows that the overall trend of inequality is negative in the aggregate measure and its subcomponents, meaning equality is improving over time. However, probably due to a series of economic crises in 1994, 1997, 2000, and 2001 in Turkey, estimated inequality index value increased between 1990 and 2000. The trend is reversed with a sharp decrease from 2000 to 2013. This might be due to policies implemented after the economic crises of the 1990s to reduce inequality in Turkey.

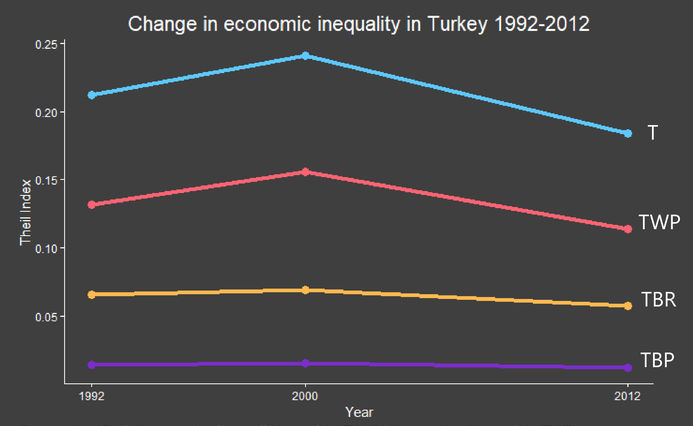

Figure 4: Evolving trends of overall inequality (T), within-province inequality (TWP), between-region inequality (TBR) and between-province inequality (TBP) in Turkey between 1992-2013 using Theil Index.

NEXT STEPS

This is an initial analysis assessing the spatial distribution of economic growth and trends in economic inequality in a country that has experienced high GDP increases over the last three decades.

In future research, I will include all years between 1992-2013 that are available in the DMSP-OLS dataset. Also using methods suggested in the literature to extrapolate missing higher resolution pixels in the DMSP-OLS dataset from the more recent and higher resolution VIIRS data, I plan to extend the analysis to more recent years after 2014.

I will also replicate the economic inequality analysis using Theil index with alternative measures at the aggregate level. Even though I might not be able to replicate the analysis for the decomposed components, I will focus on validating the aggregate inequality measure by comparing the findings from Theil index and other indices.

Coordinate System: TUREF 3 Degree GK Zone 12

Projection: Transverse Mercator

For detailed information about the data used, methods applied and results of analysis in this project please scan the QR code in the upper left side of the poster.