

Initiatives in the use of new data sources and methods in the LAC region

Giovanni Savio

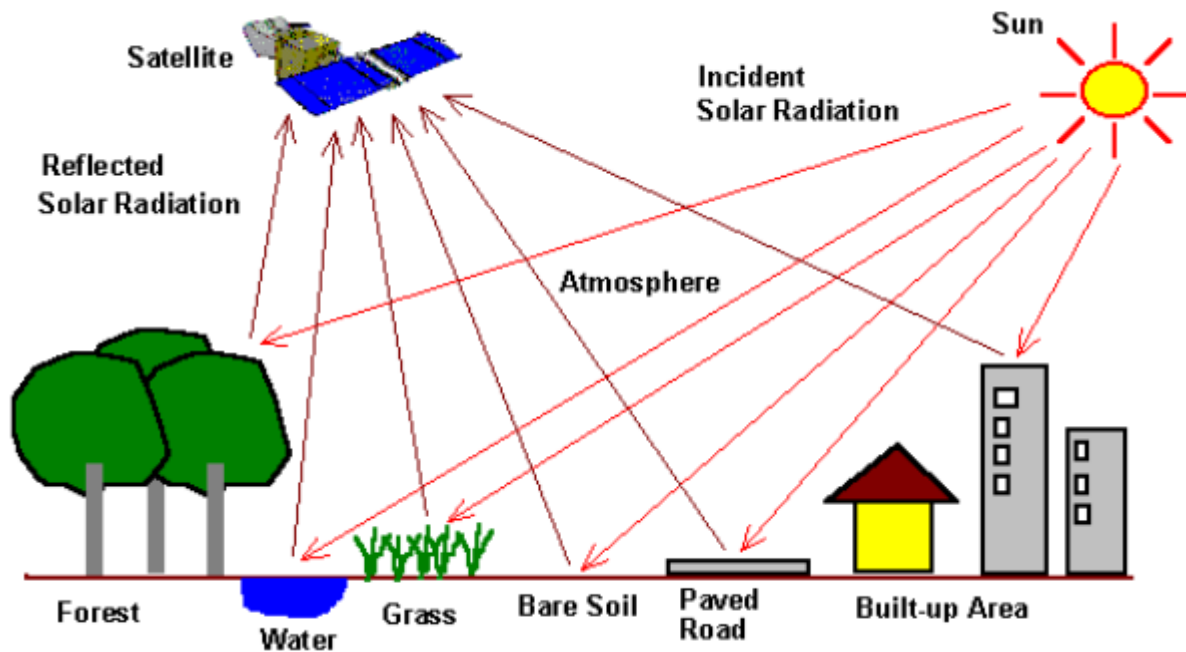
UN Economic Commission for Latin America and the Caribbean

High Level Seminar on the Future of Economic Statistics for the Arab region, UN-ESCWA- UNSD, IDB:
Riyadh, Saudi Arabia, 21-22 January 2020

Introduction

- **Objective**: To make a brief overview on use of new data sources and methods from extreme south of the world
- **Remote sensing** - How *homo statisticus sapiens* can make spatial disaggregation of GDP, poverty and other SDGs with 'observations from the above'
- **Web scraping** - Obtaining 10-15% of GDP without working too much
- **Google trend** - Why that is NOT silver bullet

Remote Sensing



Views from the above during night



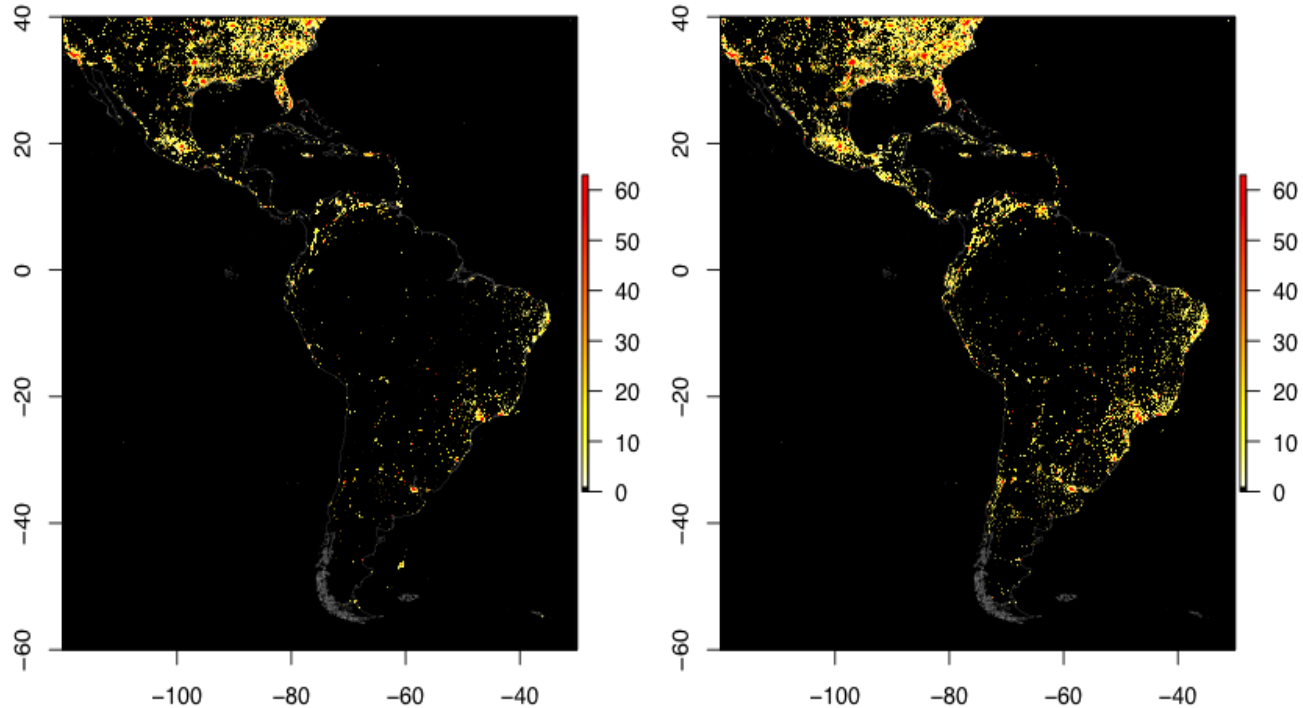
Remote Sensing During Nights

- Intensity of lights linked to:
 - a) GDP per capita, Prices, PPP (+); ECONOMIC
 - b) Poverty rates (-); SOCIAL
 - c) Population and migration flows (+); DEMOGR.
 - d) Emissions, pollution etc. (+); ENVIRON.
 - e) Others (+,-), i.e. Wars, Smuggling, Informal activities, Tourism, Urbanization

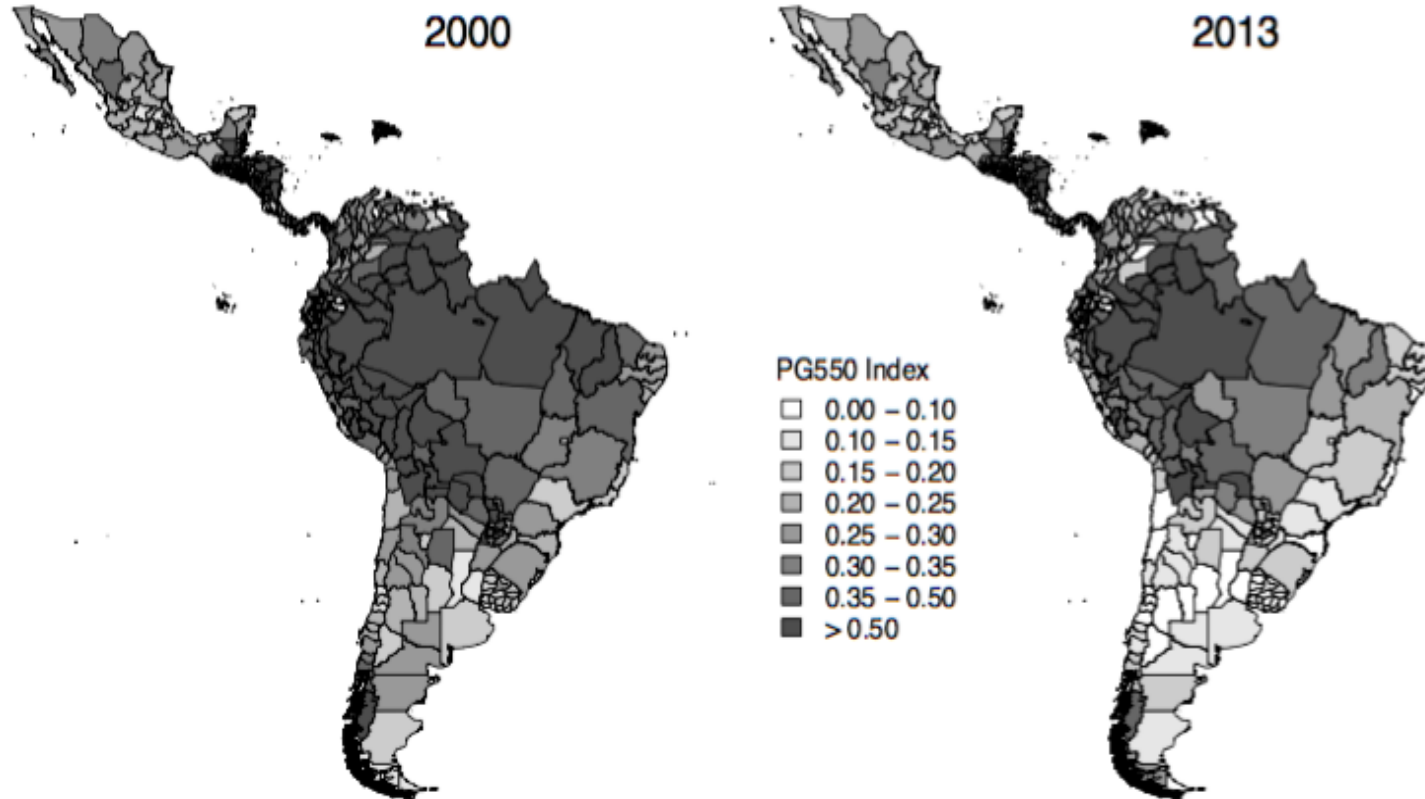
Applications of Night Lights Observations

- **With fractional panel-data models and night lights:** we obtain spatially disaggregated maps of poverty rates, in continuous time ...
- ... at virtually 1 square km ...
- ... when OFFICIAL data are available ONLY for some few scattered years, and ONLY at national level

Applications: Night Lights in LAC, 1993 (left) & 2013 (right)



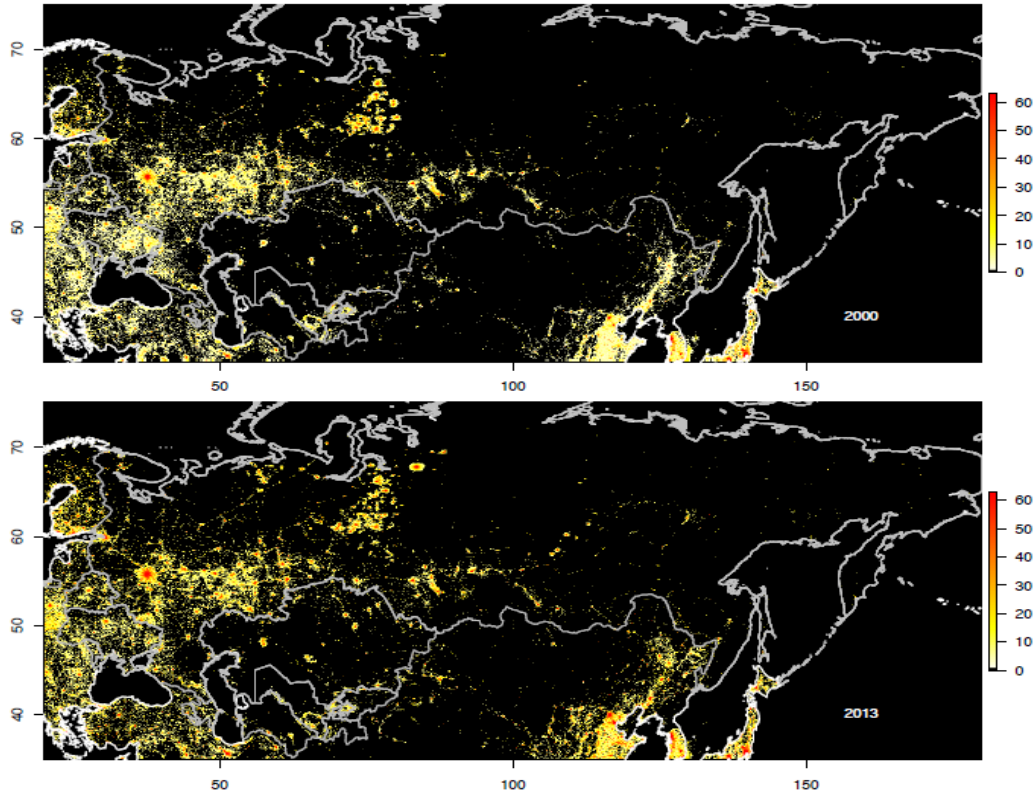
Applications: Poverty Gap in LAC



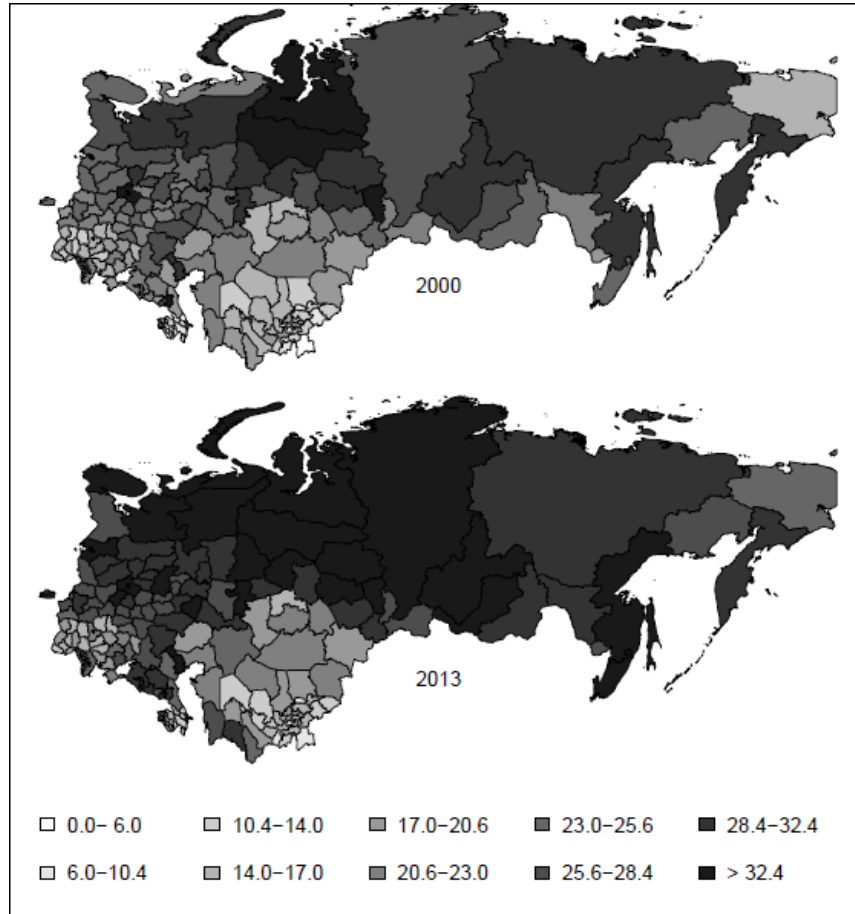
Again: Applications of Night Lights

- **With panel-data models and night lights:** we also obtain spatially disaggregated maps of GDP, PPPs and PLIs, in continuous time ...
- ... at virtually 1 square km ...
- ... when OFFICIAL GDPs sub-national data are rarely available, and PPPs and PLIs are unavailable at sub-national level

Applications: PLIs in CIS



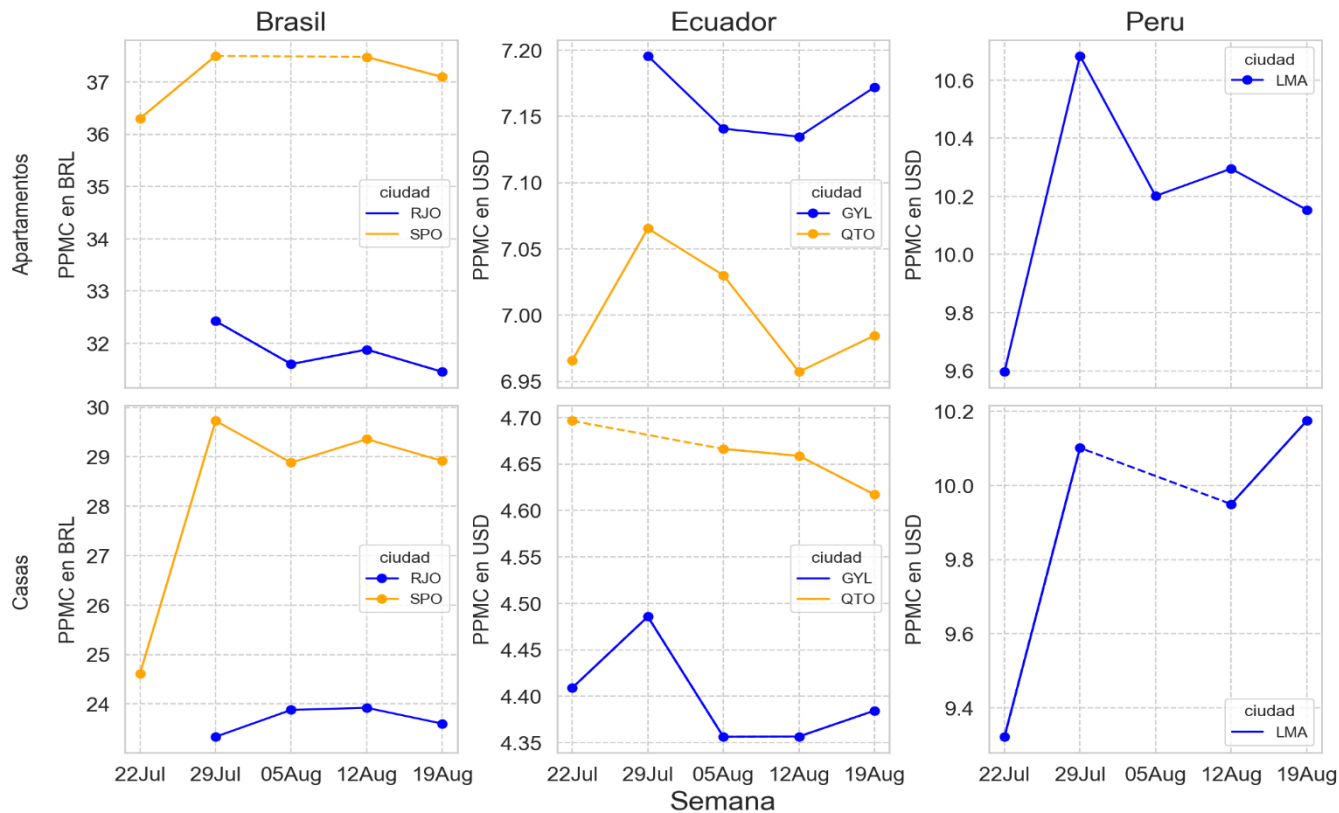
Applications: PLIs in CIS



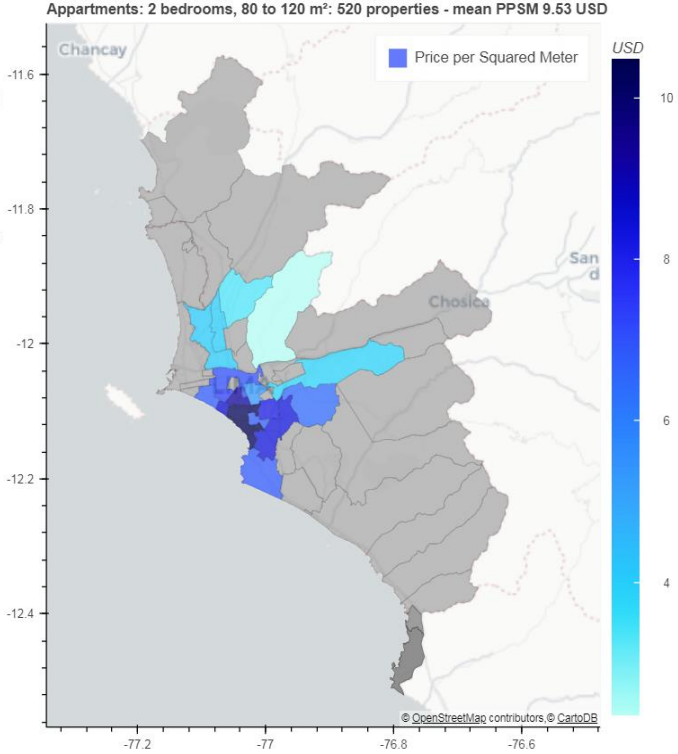
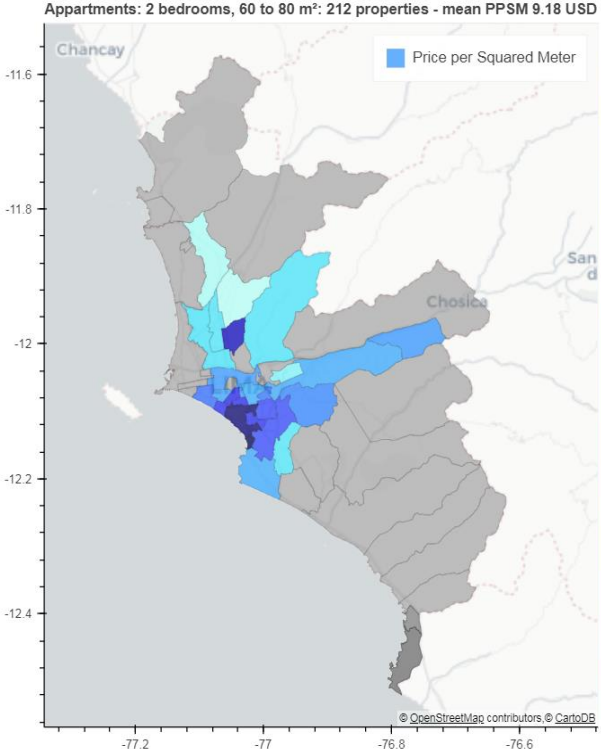
Web-scraping Application

- Objective: To obtain estimates of actual and imputed rents useful for NA and ICP purposes
- Experiment: Five cities in LA – Rio de Janeiro, São Paulo, Quito, Guayaquil and Lima
- Use of real time collection of data for 5 weeks over 13 specifications of the ICP on main rental agencies
- Use of Node.js (CBS Netherlands), Api, Geo-referencing, Google maps, Java ...

Web-scraping Application: Results



Web-scraping Application: Results (Lima)



Google Trends Data

- Used in general for forecasting/now-casting ...
- ... but CAREFULL ...
- They are not created by statisticians or for statistical purposes
- Simply represent a self-selected (non probabilistic) sample, with generating mechanisms often unknown
- Therefore, there is no guarantee that the data are representative, unless they cover the full population of interest, as it is the case for satellite remote sensing data

Conclusions

- Spatially disaggregated maps of GDP, PPPs and poverty indices, especially if updated on an annual basis, would be extremely beneficial for a number of policy-reasons
- Data obtained from remote sensing are worth considering: examples are from NASA/NOOA and EU Copernicus
- The use of VIIRS data could clearly improve on the results, permitting estimations and updating of maps at higher frequencies, but longer time series of data are necessary
- Web scraping is useful somehow, to reduce burden on official stats
...
- ... However, DO NOT BELIEVE in ALL BIG DATA sources: not all is Silver Bullet!

References

- “Sampling and modelling issues using big data in now-casting”, in *New Statistical Developments in Data Science* (eds. Verde R., Ferrari F., Petrucci A. and Racioppi F.), 2019, Springer Verlag (with M. S. Andreano, R. Benedetti, F. Piersimoni, P. Postiglione and G. Savio)
- “Mapping poverty indices for Latin American and the Caribbean countries through satellite remote sensing”, forthcoming, *Social Indicators Research*, 2020 (M. S. Andreano, R. Benedetti, F. Piersimoni and G. Savio)
- “Web scraping of internet data for house-rental services”, Paper presented at the *Eurostat-ECLAC High Level Seminar on Integrating Non-traditional Data Sources in the National Statistical Systems*, Santiago, Chile, October 1-2, 2018 (M. P. Collinao, B. Lana, R. Lara and G. Savio)
- “Mapping GDP and PPPs at sub-national level through earth observation in Eastern Europe and CIS Countries”, *Voprosy Statistiki*, 2019

THANK YOU

Giovanni Savio
giovanni.savio@un.org