Top Lights: Bright Spots and their Contribution to Economic Development.
Richard Bluhm, Melanie Krause

Dresden Presentation

Gordon Anderson

What's the paper about.

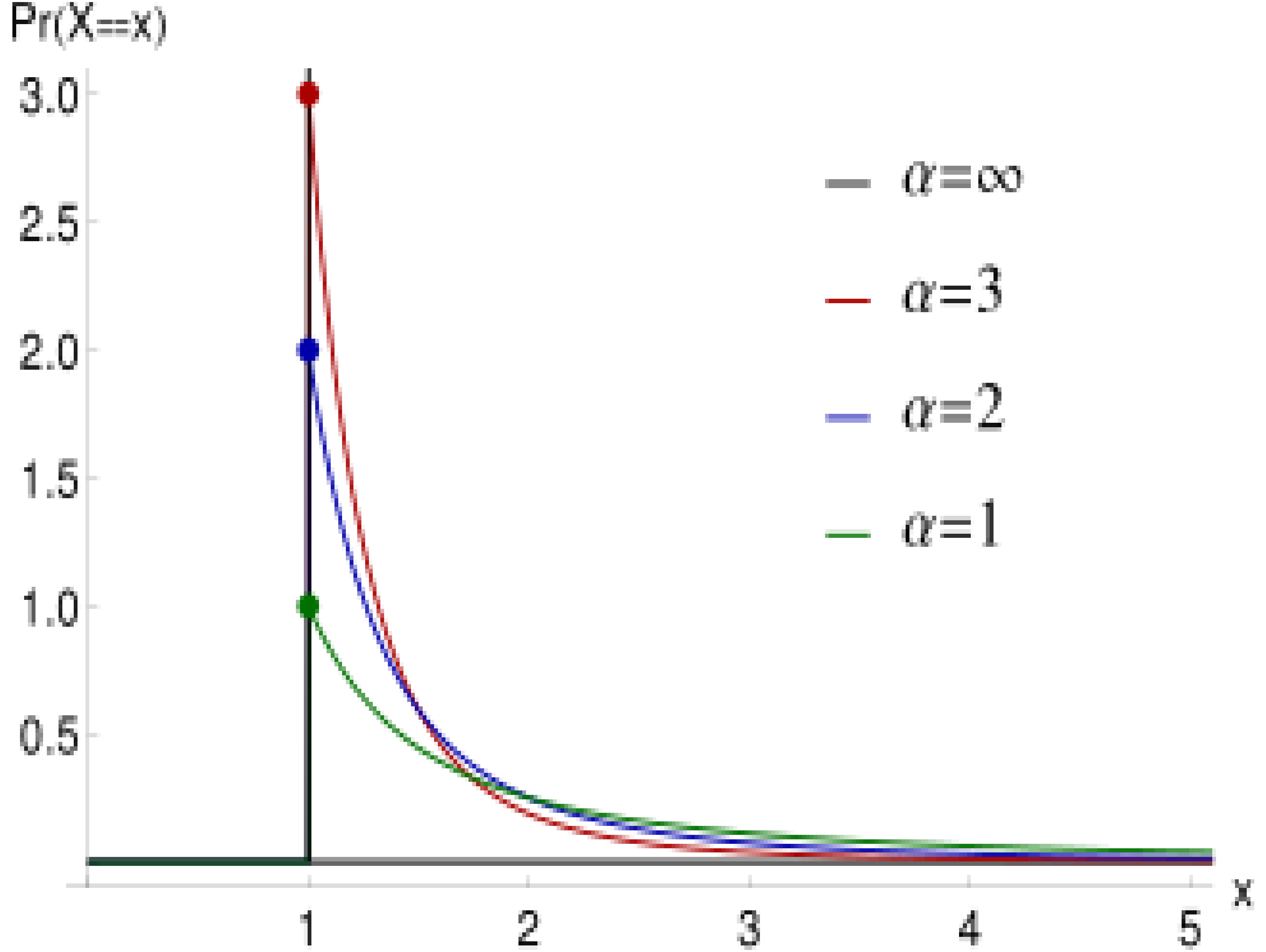
- A growing literature in economics now uses nighttime lights as a proxy for national or local economic activity.
- Advantages: publicly available as a yearly panel from 1992 onwards for nearly all parts of the world, they have a high resolution compared to regional accounts data and are measured uniformly across the globe. Hence considered comparable within and between countries, circumvent thorny discussions over exchange rate and price level adjustments.
- One drawback of this new data: they are top-coded (because of sensor saturation) in big cities and densely populated areas (i.e. the bustling center of a metropolis appears no brighter than that of a mid-sized town, distorting estimates of regional inequality and convergence).
- Contributions: Establish that top-coding matters; demonstrates that the upper tail of the distribution of light intensities at night follows a Pareto law; presents top-coding corrected estimates of average lights and spatial inequalities.

What the data are and what the problem is.

- **Satellites from the Operational Linescan System of the Defense Meteorological Satellite Program (DMSP-OLS) have orbited the earth for some decades detecting sunlit clouds. Measures of light emissions between 20:30 and 22:00 pm local time around the globe every day are a by product.**
- **Recorded data are pre-processed (cloudy day and non man made light source observations are removed) and averaged over cloud-free days resulting in images of annual light intensities from 1992 to 2013 for every pixel around the globe at a resolution of 30 by 30 arc seconds .**
- **Work with spatial random sample of 2% of all lit pixels within all countries that have a landmass larger than 500 km. A data set containing more than 4 million pixels per year from 197 countries and territories.**
- **Light intensity values are recorded in a fixed range of digital numbers (DN) from 0 “missing or dark” to 63 “bright” (can’t capture light intensity > 63 DN => the top coding problem) => rural and urban areas differences understated => understating inequality and overstating convergence measures.**

New “Radiance Calibrated” Data: A Solution to the Top Coding problem?

- Top-coding problem has received little attention so far (reliable time-series data on non-saturated lights lacking). However for seven years satellites capable of capturing the upper part of the light distribution have been flying. The resulting ‘radiance-calibrated’ data are no longer top-coded but have their own issues.
- Radiance Calibrated Data are only available for a few years
- For values that are not strictly at the top-coding boundary, large differences prevail between the two available series.
- Radiance-calibrated series varies greatly over time and is not strictly comparable across images from different satellites and years.
- Here the Radiance Calibrated Data is compared with the existing top coded data: an extension of the distribution of lights using a Pareto tail and accommodates the errors in variables problem.


Pr(X==x)
3.0
2.5
2.0
1.5
1.0
0.5
1
2
3
4
5
x
α=∞
α=3
α=2
α=1

Are Top Lights Pareto distributed? (Choice of shape parameter α is crucial)

- A rationale for a Pareto tail in the distribution of lights comes from the urban economics and top incomes literatures (Gabaix(1999) Rozenfeld et. al. (2014), Piketty (2003), Atkinson (2005), Atkinson, Piketty, and Saez (2011) and Dell (2005)).
- The Pareto CDF $F(y)=1-(y_{min}/y)^{\alpha}$ => Zipf regression of the form: $\ln(1-F(y)) = \alpha\ln(y_{min})-\alpha\ln(y) + e$
- Examines this by “eyeballing” plots and discriminant moment ratio plots (these show the coordinate pair of the coefficient of variation on the x-axis and skewness on the y-axis). Each parametric distribution has its particular curve of coordinate pairs: the plots mostly fell in the Pareto region.
- Propose correcting for top-coding by augmenting the saturated lights with a Pareto tail. Keep all of the saturated lights below the top-coding threshold replaced those above by a Pareto-distributed counterpart. Use a “rule-of-thumb” shape parameter of $\alpha = 1.75$ for the Pareto tails.

Top coded Mean and Gini

- Top-coding corrected mean luminosity μ is simply the weighted average of the bottom and top means μ_B and μ_T. The latter is the mean of a Pareto distribution starting at the top-coding threshold y_c.
- Develop a formula for GINI (Appendix A based upon Mookherjee and Shorrocks (1982)) incorporating the Subgroup Gini's using the well known Pareto distribution based Gini formula $1/(2\alpha-1)$ for the top coded group.
- The top-coding adjustment varies between 7 and 10 Gini points, with larger corrections in later years. Clearly, correcting for top-coding makes a substantial difference. Recall that the top-coding correction only affects a small fraction of lights and the remaining lights are unaltered.

Results

- Global inequality in lights started at 0.55 increased slightly over the 1990s, remained relatively stable in the first decade of the new millennium, reached a temporary trough in the aftermath of the global financial crises and great recession (2008-2010) and has since increased again to a Gini coefficient of 0.57.
- Examines specific corrections for countries.
- Outlines how to correct for top-coding in the underlying ‘stable lights’ image.

Three Prominent Research Questions.

- The first application revisits the seminal paper by Henderson et al. (2012) which established that night lights are a good proxy for GDP growth at the national level.(extend this work by examining this relationship at the subnational level in Germany).
- Concludes that, the radiance-calibrated data is indeed better suited to analyze growth in richer regions, given a lower elasticity the light-output relationship at the national level does not seem to differ systematically between rich and poor countries.
- The second application (to be completed) examines whether urban-rural differences can be quantified using nighttime lights? (expect that that top-coding will be a major issue for estimating urban-rural differences). Establish two results 1) top-coding rises with increasing population density (urbanization). Second, the economic ranking of cities using the saturated data is counterintuitive but a sensible ranking is restored after correction.
- Also intend to study regional and ethnic inequality inspired by a recent path-breaking paper by Alesina et al. (2016).

Comments: Fitting Pareto and the justification of it is a bit of a problem.

- Really neat ,I like the idea a lot. For inequality measurement accurate measurement at the top end is crucial (bigest contribution to average dif.)
- Pareto's law can be derived from Gibrat's law by invoking a reflective lower boundary on Gibrat's stochastic process. (Gabaix (1999), Anderson and Ge Journal of Regional Science and Urban Economics (2005), actually known by Operations Research guys much earlier) => distinction requires careful testing + sound theoretical reasoning for the boundary.
- Log normality or Pareto is an empirical question ~ Goodness of fit tests (Pearson or Kolmogorov Smirnov) can be applied to conditional distributions to examine this.
- Zipf regressions from $\ln(1\text{-}F(y)) = \alpha\ln(y_{min})\text{-}\alpha\ln(y)$ we get the regression: $\ln(1\text{-rank}(y_i)/n) = \alpha\ln(y_{min})\text{-}\alpha\ln(y_i) + e_i$ for i=1,n but note we may write: $(\ln(1\text{-rank}(y_i)/n)/\ln(y_{min})) = \alpha\text{-}\alpha(\ln(y_i)/\ln(y_{min})) + v_i$ so now we have a restriction i.e. intercept = regression coefficient => likelihood ratio test.
- Within this regression we can try varying the threshold to see if the shape parameter is stable (if tail is truly Pareto $f(x, \alpha, x_0) = f(x, \alpha, x_1|x>x_1)$ for all $x_1>x_0$)

Further research

- Extend satellite imagery applications.
- Neal Jean, Marshall Burke, Michael Xie, Matt Davis, David Lobell, Stefano Ermon. 2016. "Combining machine learning and satellite imagery to predict poverty". *Science*, 353(6301), 790-794.
- Argue that night-time imagery is not good at the bottom end i.e. not accurate in capturing low levels of night time bright spots.
- Use daylight images to capture features such as paved roads metal roofs etc. to measure poverty in 5 African countries via a “trained computer”…..
- What about Tucson Arizona?