Measuring Urban Economics from the Web

Samira Barzin[a,b,*], Paolo Avner[d], Jun Rentschler[d] and Neave O'Clery[a,b,c,**]

a Mathematical Institute, University of Oxford
b Oxford Martin School, University of Oxford
c The Bartlett Centre for Advanced Spatial Analysis, University College London
d The World Bank

Corresponding Authors: * samira.barzin@maths.ox.ac.uk, ** n.oclery@ucl.ac.uk

Keywords: Economic Development, Data Sparse Environments, Spatial Computation, Spatial Econometrics, Open Access Spatial Data

Extended Abstract

The world is experiencing unprecedented urban growth levels, particularly across the world's poorest regions. Currently, approx. 50% of the populations of the world's poorest countries reside in cities resulting in poor mega-cities with extremely rapid population growth (Glaeser and Henderson, 2017). This emergence of metropolises in developing countries has created immense challenges for urban governance across developing countries. The (public) transportation and road infrastructure is often observed to be at insufficient levels, resulting in low connectivity and high congestion. Further, weak institutions result, amongst others, in crime and corruption, unclear property rights lead to slums and insufficient sanitation infrastructure, and very limited financial resources represent additional crucial public policy challenges. Cities can be important drivers for economic prosperity but can equally lead to extreme poverty of many when insight and policies are inadequate. Spatial concentration of economic activity is a fundamental driver of the urban economy whereby firms benefit economically from being located in close proximity to each other

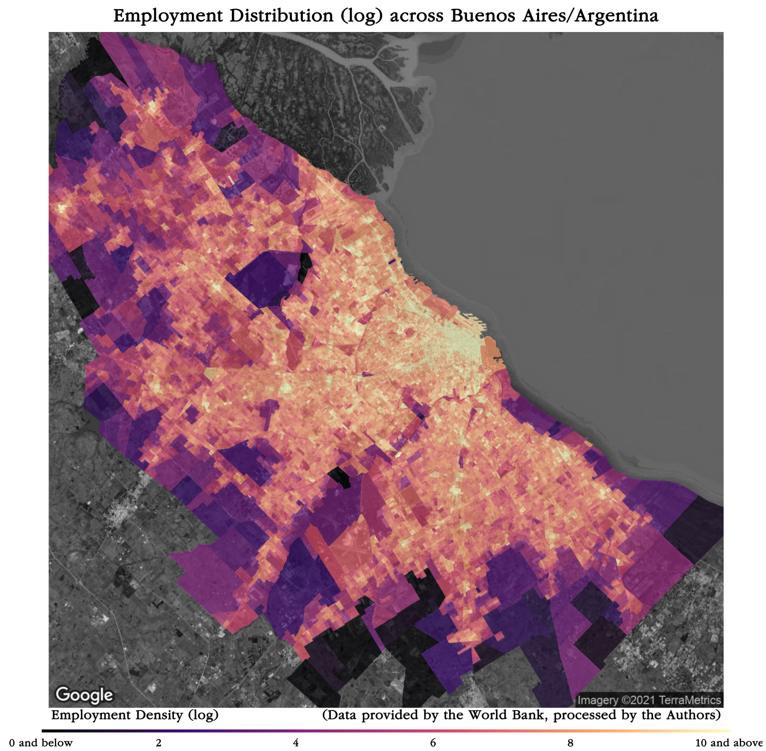

Figure 1: Distribution of Employment Over Buenos Aires

and to the consumers (agglomeration economies). Across cities in developing countries however, economic activity is often scattered throughout the city without a clear structure resulting leading to enormous economic disadvantages (Henderson *et al.*, 2016).

While there exists a solid body of literature on the urban economic dynamics of cities in developed countries, there is a crucial knowledge gap when the understanding of such dynamics is important for cities of developing countries. However, in order to address some of the above outlined urban issues, important insight is desperately needed. While there is a young literature relying on satellite imagery to gain insight into urban dynamics (see for example, Jean *et al.*, 2016), in this project, we propose to overcome this knowledge gap without the noticeable cost burden of satellite imagery by combining spatial data from multiple open access depositories. Within the analysis, we rely on the synergy of computational approaches and (spatial) data science merged with spatial econometrics to gain insight into the distribution of economic activity within cities of the developing world. We rely on a sample of 14 cities consisting of 9 located in Sub-Saharan Africa/SSA and 4 in Latin America/LAT where we have access to spatially disaggregate information of employment (see figure 1 as an example). We extract data on spatial features from the Open Street Map API on amenities and public transport, amongst others, to analyse how these features spatially correlate with employment across the sample cities. In order to strengthen the analyses, we additionally add data on various parameters obtained via the Google Earth Engine API; these include gridded population data from the WorldPop project, pollution and night lights data from NASA. Given that there exist large structural differences between cities of SSA and LAT, we split the data into separate models by geographic region for both inference and prediction. We use PCA to create an index of various road infrastructure variables, and combine this with penalised linear regressions for variable selection across all derived features. Subsequently, we include a Spatial Durbin model from the spatial econometrics methodology to explicitly exploit the spatial dependencies in the data. We further employ a generalised linear model in combination with block cross-validation and spatial random forest trees in order to train and predict employment patterns for out-of-sample cities. To account for the heterogeneity of mapping quality across cities, we derive a city level mapping quality index via accounting for the expected mapped features relative to development status of a city proxied for with night lights data in the absence of ground truth knowledge of spatial features.

The results allow us to predict the locations with the highest economic activity within cities with an accuracy of approx. 78% across the models employed; this provides us with a robust basis of the models to be used for out-of-sample predictions to identify economic activity within cities of developing countries. This insight is currently impossible when relying on traditional data sources. The algorithms produced provide an extremely important tool for policy makers to influence public investment decisions in, e.g. public transportation. Overall, this crucial insight allows cities to design better targeted policies to yield more prosperity and generate enormous social benefits, and thus aiding to lift millions of people out of poverty.

References

Glaeser, E. and J. V. Henderson, 2017. Urban Economics for the Developing World: An Introduction. Journal of Urban Economics, Vol. 98, pp. 1 - 5.

Henderson, J. V., Venables, A. J., Regan, T. and I. Samsonov, 2016. Building Functional Cities. Science, Vol. 352 (Issue 6288), pp. 946 - 947.

Jean, N., Burke, M., Xie, M., Davis, M., Lobell, D. and S. Ermon, 2016. Combining Satellite Imagery and Machine Learning to Predict Poverty. Science, Vol.353 (Issue 6301), pp. 790 - 794.