Mobility and Productivity: Quantifying Urban Economic Activity using Cell Phone Data°

Gabriel E. Kreindler [1,2] and Yuhei Miyauchi [1,2*]

[1] *Department of Economics, Massachusetts Institute of Technology, Cambridge MA*
[2] *LIRNEasia, Colombo, Sri Lanka*

How do daily commuting flows relate to the temporal and spatial distribution of urban economic activity? Simply knowing where people work will not fully capture economic activity, because locations vary in productivity. We show how the entire matrix of commuting flows obtained from cell phone data can be used to estimate productivity, which allows for a more accurate estimation of economic activity. We achieve this by estimating a gravity equation derived from a discrete choice model of job choice. To validate our approach, we use the model to predict mean residential income at commuting origins, which we compare with a new dataset of nighttime lights. We discuss potential applications.

Human mobility and economic activity are intertwined. Previous work has established that international migration flows are responsive to economic activity in the destination country, as well as to distance [1], [2]. Regional and seasonal migration also respond in a similar fashion [3], [4]. These results are suggestive of a relationship between mobility and economic activity at much finer temporal and geographic scales. However, to date this relationship has not been studied quantitatively mainly due to data limitations. This is the goal of our project. We study the link between daily mobility (commuting) flows and economic activity within cities, and quantify it using cell phone data.

Our project has three contributions. First, we set up a theoretical model that links individual home- and job-choice decisions, commuting flows and economic productivity, and show that it can be implemented empirically through a gravity model. Second, we use cell phone data from Sri Lanka to construct a measure of commuting with fine urban spatial and temporal variation, and proceed to estimate the model.[1] Third, we validate the model's performance using a new and improved data source of nighttime lights.

Our work builds on a recent literature on urban sensing [5], and the measure of economic activity may eventually assist policy in urban areas. This detailed data can help improve economic policy [6], as well as benefit urban-planning by making it more responsive to detailed trends in the concentration of economic activity.

Discrete Choice Model and Gravity Equation

In the model, based on [7], workers choose their work destination by trading off the wage offered at each destination, against the distance needed to reach each destination. Naturally, high wages may compensate for the cost of commuting. Other idiosyncratic factors that influence job choice are modelled as random shocks from a pre-specified distribution (Fréchet or type II extreme value). This setup implies that the commuting probability π_{ij} that a worker residing in origin i commutes to destination j satisfies the following gravity equation:

$$\log(\pi_{ij}) = \psi_j + \beta \log(D_{ij}) - \mu_i + \varepsilon_{ij}, \qquad (1)$$

where ψ_j is a log transformation of the wage at j, D_{ij} is a measure of distance between i and j, μ_i captures origin-specific factors, and ε_{ij} is measurement error. Intuitively, ψ_j captures a destination's attractiveness, after controlling for distances to all origin locations, and their respective sizes. The benefit of using an explicit model of workers' decisions is that it allows us to back out other useful economic measures, such as the output and the average residential income at a particular location.

In the full paper, we study several features and extensions of the model, such as the importance of the aggregation level, (i, j) pair specific factors, and worker heterogeneity. The model can also accommodate a variety of assumptions on how workers choose their home location.

Commuting Flows and Model Estimation

Empirically, the model can be mapped to the real world using commuting flows extracted from cell phone data. We use a simple algorithm to construct daily commuting flows from CDR data, building on the literature on this topic [8], [9]. Figure 1 shows the distribution of smoothed commuting flows from a particular origin location (i.e. a given cell tower Voronoi cell) in the greater metropolitan

° The authors are grateful to the LIRNEasia organization for providing access to cell phone data and an excellent working environment, to Sriganesh Lokanathan, research manager at LIRNEasia, whose dedication and relentless efforts made this project possible, and to Danaja Maldeniya for constructive feedback on the project. We acknowledge funding from the International Development Research Centre (IDRC).

* Contact: gek@mit.edu; miyauchi@mit.edu.

[1] Our approach will perform significantly better in *urban* areas, due to higher cell tower density relative to less developed areas.

Colombo area. What is noteworthy is that there is significant variation in commuting flows even after accounting for commuting distance, and this variation appears to capture anecdotal patterns of commuting in the Colombo area.

We estimate equation (1) on commuting volumes between pairs of towers in Sri Lanka. There are 3,047 towers and $\approx$ 1,7 million pairs of towers in our sample, covering over 300 million commuting trips. We use a linear regression model with two sets of fixed effects (corresponding to origin and destination locations). Using the estimated fixed effects and other coefficients, we construct the residential income and output measures.

Comparison with Nighttime Lights

Having estimated economic productivity (wages) and economic activity (output), exclusively from mobility flows derived from cell phone data, we would ideally proceed to validate these measure using independent wage and output measures. Unfortunately, in our context this type of data is only available aggregated at province level.

Instead, we present results from a validation exercise using a measure Φ_i of mean *residential* income in origin location i, computed using the model. Intuitively, Φ_i captures the average of wages brought "home" by workers who live in i and who work in various other destinations. To assess whether Φ_i contains non-trivial information, we compare it with nighttime lights, which is a recognized measure of residential income [10]–[12]. We use a new version captured by the VIIRS satellites and curated by the Earth Observatory Group (EOG) at NOAA. The VIIRS data has higher spatial and temporal resolution than the older OLS data, and it does not have a saturation point.

Figure 2 shows a graphical comparison: nighttime light VIIRS in the top panel, and the $\log(\Phi_i)$ measure (at the tower cell level) in the bottom panel. The correspondence is generally good, yet there are clear points where the model can be improved (e.g. patterns along roads). In the paper we show using regression analysis that the income measure is informative after controlling for population density (interpolated from the census), and various simple indicators derived from cell phone data.

Ongoing and Future Work

We plan to exploit the time variation in the cell phone data to look at changes in the spatial distribution of economic activity over short and medium time horizons. For example, we can investigate the effects of short term disturbances, such as the transportation restrictions imposed by national events. We can also study whether, in the medium term, economic activity becomes more concentrated around city centers, or whether it actually moves towards city outskirts.

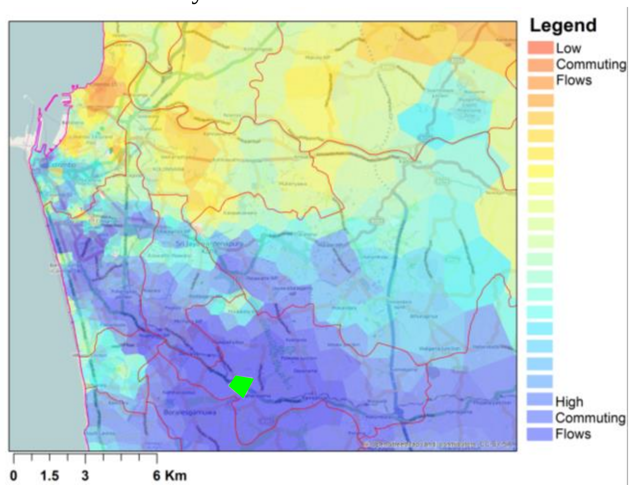

Figure 1. Smoothed commuting flows from a given origin tower cell (indicated by the green cell). Red lines indicate sub-district (divisional secretariat) boundaries.

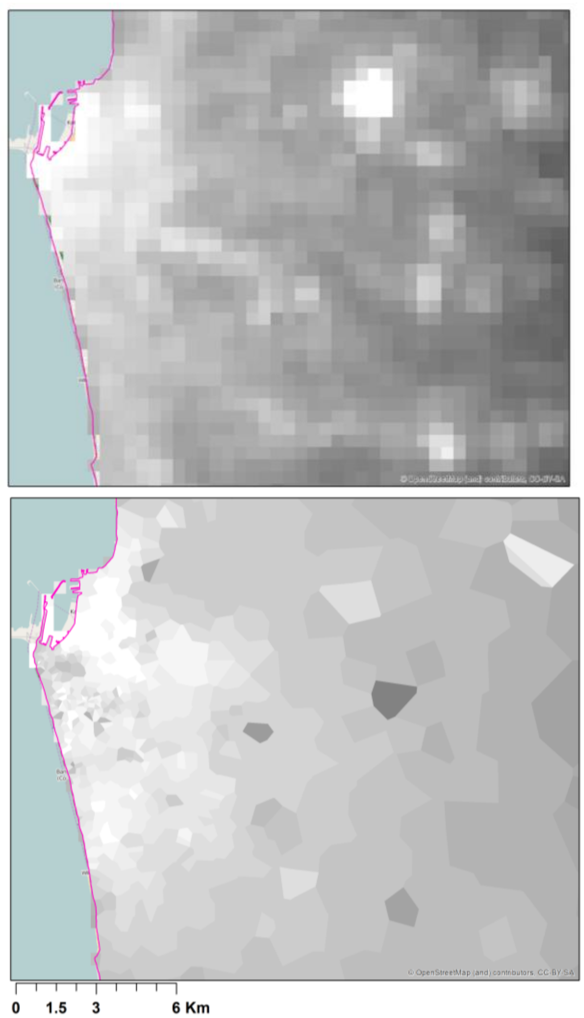

Figure 2. The top panel shows a log transformation of VIIRS nighttime lights. (The bright spot corresponds to an oil refinery.) The lower panel shows the measure of residential income $\log(\Phi_i)$ derived from the model (white corresponds to higher values).

References

[1] A. M. Mayda, "International migration: A panel data analysis of the determinants of bilateral flows," *J. Popul. Econ.*, vol. 23, pp. 1249–1274, 2010.

[2] F. Ortega and G. Peri, "The Effect of Income and Immigration Policies on International Migration," *NBER Work. Pap. Ser.*, no. 18322, 2012.

[3] O. J. Blanchard and F. Katz, "Regional Evolutions," *Brookings Pap. Econ. Act.*, vol. 23, no. 1, pp. 1–76, 1992.

[4] D. Hare, "'Push' versus 'pull' factors in migration outflows and returns: Determinants of migration status and spell duration among China's rural population," *J. Dev. Stud.*, vol. 35, no. January 2015, pp. 45–72, 1999.

[5] F. Calabrese, L. Ferrari, and V. D. Blondel, "Urban Sensing Using Mobile Phone Network Data: A Survey of Research," *ACM Comput. Surv.*, vol. 47, no. 2, pp. 1–20, Nov. 2014.

[6] J. Mitchell, R. J. Smith, M. R. Weale, S. Wright, and E. L. Salazar, "An indicator of monthly GDP and an early estimate of quarterly GDP growth," *Econ. J.*, vol. 115, pp. 108–129, 2005.

[7] G. M. Ahlfeldt, S. J. Redding, D. M. Sturm, C. Fieler, G. Grossman, B. Honore, U. Mueller, S. Kortum, E. Morales, and F. Hall, "THE ECONOMICS OF DENSITY :," *NBER Work. Pap. Ser.*, no. 20354, 2014.

[8] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating Origin-Destination Flows Using Mobile Phone Location Data," *IEEE Pervasive Comput.*, vol. 10, no. 4, pp. 36–44, Apr. 2011.

[9] P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, and M. C. González, "Understanding road usage patterns in urban areas.," *Sci. Rep.*, vol. 2, p. 1001, Jan. 2012.

[10] J. V. Henderson and A. Storeygard, "Measuring Economic Growth from Outer Space," *Am. Econ. Rev.*, vol. 102, no. 2, pp. 994–1028, 2012.

[11] X. Chen and W. D. Nordhaus, "Using luminosity data as a proxy for economic statistics.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 21, pp. 8589–8594, 2011.

[12] C. Mellander, K. Stolarick, Z. Matheson, and J. Lobo, "Night-Time Light Data : A Good Proxy Measure for Economic Activity ?," no. 315, p. 33, 2013.