Final report

International Growth Centre

Capturing urban mobility with cell phone data with an application of flyover opening in Dhaka

Yuhei Miyauchi

February 2017

When citing this paper, please use the title and the following reference number:
C-31307-BGD-2

DIRECTED BY

FUNDED BY

Capturing Urban Mobility with Cell Phone Data with an Application of Flyover Opening in Dhaka

Yuhei Miyauchi[1]

February 2017

Abstract

In this project, I explore the usage of cell phone data for urban transportation planning in developing countries. I lay out a methodology to extract urban trip patterns of people from cell phone data at the spatially and temporarily fine scale. The method is applied to Dhaka, Bangladesh and validated through the existing random-sample travel survey. To illustrate one example of the usage of such data, I analyze how the opening of Jatrabari-Gulistan Flyover in October, 2013 has changed the urban trip patterns. I find that about 23.8% more trips are generated on the routes that crosses over flyover relative to other routes. There is a within-day variation in the generated trips; in the mornings more trips are generated from suburban area (Narayangonj area) toward the central business district (Motijheel area), and the other way around in the evenings.

[1] Miyauchi: MIT, miyauchi@mit.edu. This report is prepared as a final report for IGC Small Grant Project: "Impact of Traffic Infrastructure on Urban Land Use in Dhaka." I am grateful to grateful to Ryosuke Shibasaki for navigating us through the cell phone data used in this study forin Bangladesh, as well as. Anisur Rahman and Takashi Hiramatsu for the access to the DHUTS data in Bangladesh. I sincerely appreciate support and guidance of David Atkin, Abhijit Banerjee, Arnaud Costinot and Esther Duflo, and the comments from Gabriel Kreindler. I am also grateful to Akib Khan and Akira Matsushita for their great research assistantship, and to IGC Bangladesh team for supporting this project both financially and substantially. I also appreciate the Nakajima Foundation for the financial support for my Ph.D scholarship.

1 Introduction

Urban transportation planning is a big pain in the neck of policy makers in developing countries. Virtually all cities with rapid economic growth are facing severe traffic congestion and urban transportation problems. What makes the situation particularly difficult is the lack of up-to-date information; precisely because they are rapidly growing, they do not have up-to-date data that they can rely their policies on. Traditional approaches in developed countries, particularly the usage of detailed travel survey, is costly, and frequently conducting them to keep the information up-to-date is difficult.

This project aims to explore the usage of cell phone data for urban transportation planning. Cell phone data, or more precisely call detail records (CDRs), is the collection of timestamps and locations of all customers automatically collected by cell phone operators. High penetration of cell phone data, as well as its near real-time nature, opens up the scope of costless resource for urban transportation planning.

In this report, I conduct two exercises. First, use data on urban trips of people extracted from cell phone data in Dhaka, Bangladesh, and validate it with the existing random-sample travel survey. Second, I analyze how the opening of Jatrabari-Gulistan Flyover has changed the urban trip patterns. Jatrabari-Gulistan Flyover is a long-awaited transport infrastructure in Dhaka constructed in the traffic bottleneck between the central business district (CBD; Motijheel area) and suburban area (Narayangonj area). By comparing the month before and after the construction of the flyover, I show that about 20% more trips are generated across the flyover relative to other origin-destination pairs. Furthermore, I find more increase of trips from suburban area to the CBD in the mornings and the other way around in the evenings.

There is some existing literature that utilizes CDR data to construct urban mobility patterns in both developed (e.g. Calabrese et al. (2011), Wang et al. (2012) in the US) and developing countries (e.g. Iqbal et al. (2014), Hasegawa et al. (2014) in Bangladesh, Demissie et al. (2016) in Senegal, Kreindler and Miyauchi (2016) in Sri Lanka). These data are validated with census-tract level work-home locations (Wang et al. (2012)) and traffic volume at the intersection level (Iqbal et al. (2014), Hasegawa et al. (2014)). This project contributes to this literature by validating the mobility data from CDR by random-sample travel survey, which reports the diary of trips made by the respondents at different time of the day. Such information has richer within-day trip information than just using work-home commuting flows or traffic volumes.

There is also some emerging literature on the usage of CDR data to analyze the impact of traffic infrastructure. For example, Kujala et al. (2016) analyze the impact of inter-city highways on typical travel time using the opening of the Dakar-Diamniadio highway in Senegal. In addition to the fact that they mainly focus on travel time rather than travel volume, my setting is different from them in that the flyover studied in this project is in the center of the city and affects urban mobility rather than inter-city trips.

The remaining section of the report is organized as follows. In Section 2, I briefly describe the data sources that are used in this study. In Section 3, I discuss how I construct the trips data from cell phone data and validate it using random-sample travel survey. In Section 4, I study the impact of Jatrabari-Gulistan Flyover. Section 5 concludes and discusses the future research.

2 Data Sources

In this section, I describe the cell phone data and the method to construct commuting flows from it. I also describe other data sources that I use for complementary analysis in the subsequent sections.

2.1 Trip data from Call Detail Record (CDR)

The data is prepared by Asian Development Bank for the project (A-8074REG: “Applying Remote Sensing Technology in River Basin Management”), a joint initiative between ADB and the University of Tokyo. The original data includes observations of every single phone call made by all the cell phone holder. Each observation contains a timestamp, the user identifiers of the participants, and the cell antennas to which they are connected. The data covers seven million people in Greater Dhaka for the period of 1st August 2013 and 31st January 2014, except for the month of October 2013. There are about 2000 cell tower locations in Greater Dhaka area. This roughly accounts for squared kilometer per cell tower on average, but the density of tower has high variation across the city. For the purpose of the study, I use the trips data extracted from the original CDR data in this study.

Particularly important point about CDR to capture trip patterns is the fact that CDR is a sparse dataset; I only observe the location of cell phone holders *when they make a transaction*. It is therefore important to validate the trip data from CDR with true information of trips to see whether the data sparseness causes substantial sample selection bias or not.

2.2 DHUTS Travel Survey

To validate my cell-phone based trip data, I will utilize the raw data of travel survey conducted in Dhaka Urban Transport Network Development Study (DHUTS) by Japan International Corporation Agency (Japan International Cooperation Agency (2010)). The travel survey is conducted to formulate the Urban Transport Network Development Plan integrated with urban development plan of DMA for the period up to 2025, and to draw general outline of the urban transport projects to be implemented on the priority basis. The survey covers about 18,000 households randomly extracted based on the residential location from the voter registry.

Among the several components of the survey, I will mainly utilize the travel diary. The travel diary asks respondents to report all the trips that they have made for a particular day, and it includes the information about the starting time and location, ending time and location, mode of transportation, and purpose of the trips. This information allows us to validate the trip information extracted from the cell phone data at temporarily and spatially fine scale.

DHUTS travel diary geocodes the origin and destination locations within Dhaka City into 90 Traffic Analysis Zones (TAZs). When I validate the trips from CDR data, I assign cell phone towers at this level and define the trips at this spatial unit.

2.3 Google Map Directions API Data

We use Google Map Directions API data to proxy pairs of locations which are directly affected by the Jatrabari-Gulistan Flyover. I query Google Map Directions API for the shortest travel routes, and check whether the route involves a part of Jatrabari-Gulistan

Flyover. I define that the route from A to B involves the usage of flyover if either A to B or B to A involves the usage of the flyover.[2]

Usage of Google data as a source of urban planning in developing countries is also becoming popular. In particular, Google Traffic API provides a real-time traffic speed wherever available.[3] Unfortunately Google *Traffic* is not yet available in Dhaka as of February 2017. Google *Map* is available but the estimated arrival time is not accurate because it is not based on actual traffic data.

Although the Google Traffic data is an extremely useful data source for capturing real-time traffic *speed*, it does not provide the trip volume information (how many people travel from where to where at what time). Therefore, even after Google Traffic becomes available in Dhaka (which I hope it will soon), I believe the CDR data will remain useful data source for urban transportation planning.

3 Constructing and Validating Trips from Cell Phone Data

In this section, I describe the algorithm of extracting trips from CDR data, and validate the data with DHUTS travel survey. The main finding is that although the two measures are extremely well-correlated, there is some systematic quantitative bias. More analysis should be conducted to reduce the bias in future work.

3.1 Trip Data from CDR

In this report, I define trips in the following simple way: For a specific day, hour, and cell tower location, I define number of trips from location A to location B to be the number of cell phone users that accessed in both towers within 3 hours.[4]

This naïve approach has at least one shortcoming: It is likely to overestimate the long-distance trip. For one thing, if one compares the person who makes a trip for 1 hour and another with 2 hours, s/he stays in the origin and destination location longer than the latter (within 3 hour window, the first person stays in origin and destination location for 2 hours but the latter stays only 1 hour). Hence, the probability of observing long-distance trip is lower. Moreover, for longer-distance trips, the probability that a person makes a phone call on the way is higher (simply because s/he spends more time for the trip), which underestimate the trip distance of this person. I in fact confirm that such bias exists in the next section.

3.2 Validation of CDR Trip Data with DHUTS

To validate the CDR trip data, I compare the probability of trips to different destination locations conditional on origin locations. The origin and destination locations are defined at

[2] In practice, we query only a subset of cell-tower pair on Google Map API (about 30,000 pairs), and evaluated the usage of flyover for the rest of the tower pairs by spatial interpolation. There are nearly 2 million cell tower pairs with positive trip data, and querying all of them on Google Map is not realistic.

[3] For example, Hanna et al. (2016) uses real-time traffic information from Google Traffic API to analyze the travel speed and the implication of congestion reduction policy.

[4] Construction of trip data from original CDR data is conducted by Asian Development Bank for the project (A-8074REG: “Applying Remote Sensding Technology in River Basin Management”), a joint initiative between ADB and the University of Tokyo

90 TAZs in DHUTS, and I take four different time windows (5:00-9:59, 10:00-14:59, 15:00-19:59, 20:00-23:59) and define the sample at origin-destination-time tuples.[5]

Panel (A) of Figure 1 provides a direct comparison of these two measures. The two measures are extremely well correlated: A simple regression has R-squared of 0.289. At the same time, there is some bias quantitatively: The slope is shallower than the 45 degree line in red. In other words, cell phone trip data underestimates the probability of trip for small-probability origin-destination pairs, while it overestimates the probability of trip for large-probability origin-destination pairs.

Panel (B) of Figure 1 provides a potential explanation of this bias. The graph plots the probability of trips against the geodesic distance between the origin and destination for the two trip data. The cell phone trip data shows steeper slope in distance; implying that it underestimates the trip volume of distant trips and overestimate close trips. This is consistent with the speculation explained in Section 3.1.

In sum, although the cell phone trip data is well-correlated with DHUTS data, there is some caution in interpreting the number quantitatively. Improving on this point is left for future research.

4 Impact of Jatrabari-Gulistan Flyover

In this section, I show that Jatrabari-Gulistan Flyover affects within-city travel patterns. After briefly explaining the background of Jatrabari-Gulistan Flyover, I show that the Jatrabari-Gulistan Flyover increased the trip volume by 23.8% relative to non-flyover routes measured by cell phone trip data. I further argue that there is substantial time heterogeneity in impact: In the mornings more trips are generated from suburban area (Narayangonj area) toward the CBD (Motijheel area), and the other way around in the evenings.

4.1 Background of Jatrabari-Gulistan Flyover

Jatrabari-Gulistan Flyover, also known as Mayor Mohammad Hanif Flyover, is a 11.8km-long flyover opened on October 11, 2013. It extends from Jatrabari to Sayedabad, Tikatuli, Joykali Mondir, Kaptanbazar and Gulistan. In particular, it serves as a connecter to CBD (Motijheel area) and suburban area (Narayangonj area). The construction started on June 22, 2010 as the largest public-private partnership investment in Bangladesh.

Since the cell phone data covers from August 2013 to January 2014, it allows us to analyze the trip pattern change before and after the opening of the flyover. Unfortunately the data does not cover October 2013, and I cannot analyze the precise date of the opening of the flyover.

4.2 Impact of Jatrabari-Gulistan Flyover on Trip Volumes

Figure 2 shows the heat map of the sum of commuting flows from and to the CBD (Motijheel area) in September (Panel (A)) and the growth rate from September to November (Panel (B)). The flyover is marked in black, and the CBD is marked with green. One can visually recognize from Panel (B) that the south-eastern part of the flyover, suburban area in the other side of CBD, exhibits increased trip volume.

[5] Since the trips in both datasets from 0:00-4:59 are relatively rare, I omit the trips in this time interval from the analysis.

The increased trip volume by the flyover is confirmed formally by the regression. Table 1 shows the regression results of the impact of the flyover. Columns (1) to (4) show the impact on the growth rate of trip volume from September to November 2013. In all regressions I control for the baseline volume fixed effects. In the baseline results (Column 1), the trip volume increased 23.8% on the flyover routes relative to non-flyover routes. The pattern still remains even if I control for origin cell tower fixed effects (Column 2) and destination cell tower fixed effects (Column 3).

In addition to the trip volume, I also find that the minimum trip duration between cell towers is decreased on average on flyover routes relative to non-flyover routes. The minimum trip duration is defined by the minimum trip duration for each cell tower within a month. Column 5 shows that the flyover route reduced the minimum trip duration 4 minutes more relative to non-flyover routes. Although this is an encouraging results, I keep the results only as suggestive, as this minimum trip duration data is not validated in the literature or in this work. In particular, the sparseness of the CDR data (i.e. I observe the location of people only when they make a phone call) makes the interpretation particularly difficult.

4.3 Heterogeneous Impact of Jatrabari-Gulistan Flyover within a Day

The impact of flyover is heterogeneous within a day. To see this, I look at the heterogeneous impact of the flyover stratified by four different time windows (5:00-9:59, 10:00-14:59, 15:00-19:59, 20:00-23:59) and the direction of the flyover routes (westward or eastward; westward means from Narayangonj side to the CBD side).

I first discuss how the distribution of trips on flyover routes look by time and direction. Panel (A) of Figure 3 shows the average trip volume for each time of the day and by the direction for flyover routes. One can recognize that the trip volume is relatively large for westward than eastward in the mornings, and the other way around in the evenings. This is suggestive of the fact that there is net positive commuting from Narayangonj side to the CBD side during the day.

Panel (B) of Figure 3 plots the flyover-impact regression coefficients of the interaction of the time window and the direction. The regression specification is the same as in Table 1. I find that the impact on eastward trips is larger in the mornings but the other way around in the evenings. This is suggestive of more commuting trips during the day are generated from Narayanganj to the CBD.

5 Conclusion and Future Work

In this project, I explore the usage of cell phone data for urban transportation planning in developing countries. I first lay out the methodology to extract urban trip patterns of people from cell phone data at the spatially and temporarily fine scale. The method is applied to Dhaka, Bangladesh and validated through the existing random-sample travel survey. Although the cell phone trip data is well-correlated with DHUTS data, there is still some quantitative bias. Next, I analyzed how the opening of Jatrabari-Gulistan Flyover has changed the urban trip patterns. I find that about 23.8% more trips are generated across the flyover relative to other origin-destination pairs. There is within-day variation in the generated trips; in the mornings more trips are generated from suburban area (Narayangonj area) toward the central business district (Motijheel area), and the other way around in the evenings.

There are several future directions of the research. First, the method of extracting trips should be refined to better reflect the true trip volumes quantitatively. Some methodological refinement should be considered to address the speculation discussed in Section 3.1. Second, more detailed analysis on the Jatrabari-Gulistan flyover should be done. In particular, the nature of cell phone trip data as behavioral decisions (when to leave to where) should be exploited to understand the equilibrium interactions of traffic (how the flyover affects traffic in other areas) as is suggested by Hanna et al. (2016), or infer the impact on economic activity (Kreindler and Miyauchi (2016)).

References

Calabrese, F. et al., (2011). “Estimating Origin-Destination Flows Using Mobile Phone Location Data.” *IEEE Pervasive Computing*, Vol.10, No.4, pp.36–44.

Demissie, M.G. et al., (2016). “Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal.” *IEEE Transactions on Intelligent Transportation Systems*, Vol.PP, No.99, pp.1–0. Available at: http://ieeexplore.ieee.org/ielx7/6979/4358928/07416225.pdf?tp=&arnumber=7416225&isnumber=4358928%5Cnhttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7416225.

Hanna, R., Kreindler, G. and Olken, B.A., (2016). “CITYWIDE EFFECTS OF HIGH-OCCUPANCY VEHICLE RESTRICTIONS: EVIDENCE FROM THE ELIMINATION OF ‘ 3-IN-1 ’ IN JAKARTA.” *mimeo*.

Hasegawa, Y. et al., (2014). “Transportation Melting Pot Dhaka: Road-link Based Traffic Volume Estimation from Sparse CDR Data.” In *Proceedings of the First International Conference on IoT in Urban Space*.

Iqbal, M.S. et al., (2014). “Development of origin-destination matrices using mobile phone call data.” *Transportation Research Part C: Emerging Technologies*, Vol.40, pp.63–74. Available at: http://dx.doi.org/10.1016/j.trc.2014.01.002.

Japan International Cooperation Agency, (2010). *Preparatory survey report on Dhaka urban transport network development study (DHUTS) in Bangladesh : final report.*, Available at: http://open_jicareport.jica.go.jp/710/710/710_101_11996774.html.

Kreindler, G. and Miyauchi, Y., (2016). “Quantifying Commuting and Urban Economic Activity using Cell Phone Data.” *mimeo*.

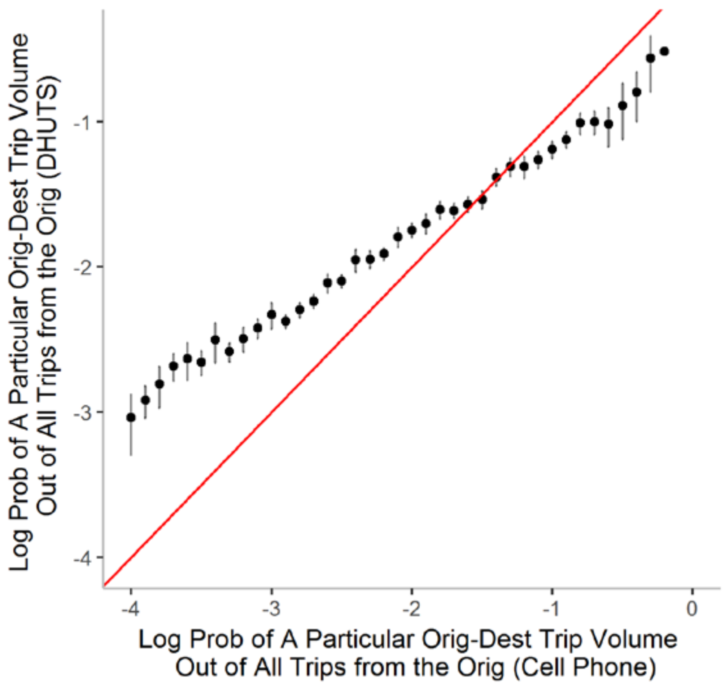
Kujala, R., Aledavood, T. and Saramäki, J., (2016). “Estimation and monitoring of city-to-city travel times using call detail records.” *EPJ Data Science*, Vol.5, No.1. Available at: http://dx.doi.org/10.1140/epjds/s13688-016-0067-3.

Wang, P. et al., (2012). “Understanding road usage patterns in urban areas.” *Scientific reports*, Vol.2, p.1001. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3526957&tool=pmcentrez&rendertype=abstract [Accessed November 8, 2013].

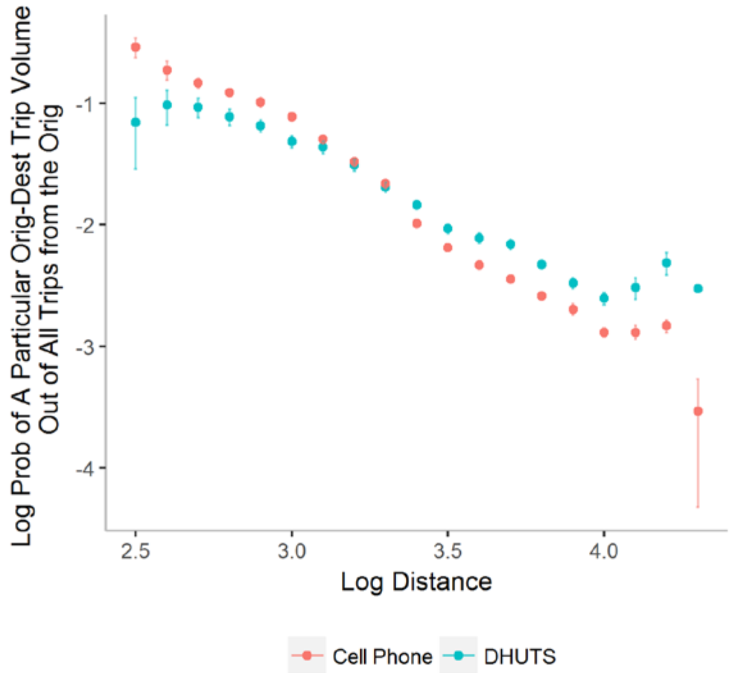
Figures and Tables

Figure 1. Validation of Trips from Cell Phone Data.

(A) Comparing Trip Patterns between Cell Phone Trip data and DHUTS



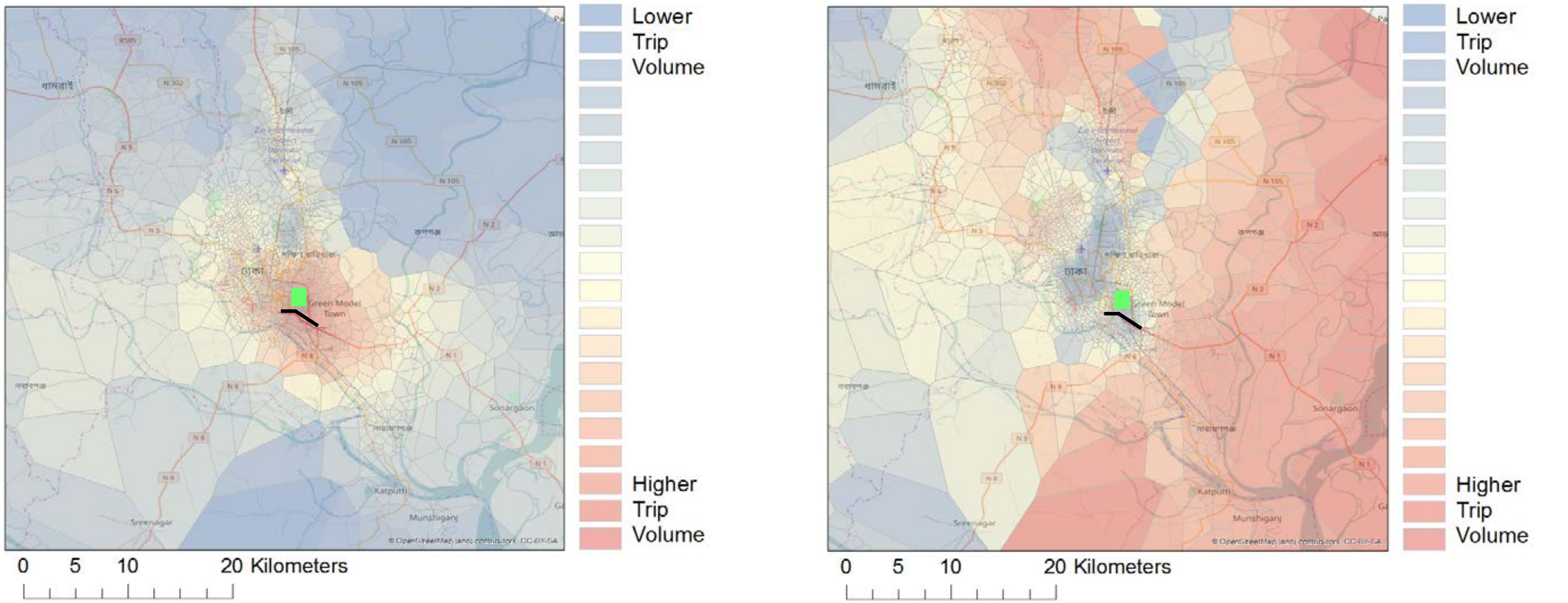
(B) Trip Probability as a Function of Distance



Note: The origin and destination locations are defined at 90 TAZs in DHUTS, and I take four different time windows (5:00-9:59, 10:00-14:59, 15:00-19:59, 20:00-23:59) and define the sample at origin-destination-time tuples. Dots are the means and the thin lines are the 95% confidence intervals. The red line in Panel (A) is the 45 degree line.

Figure 2. Change of Trip Pattern before and after Jatrabari-Gulistan Flyover Opening.

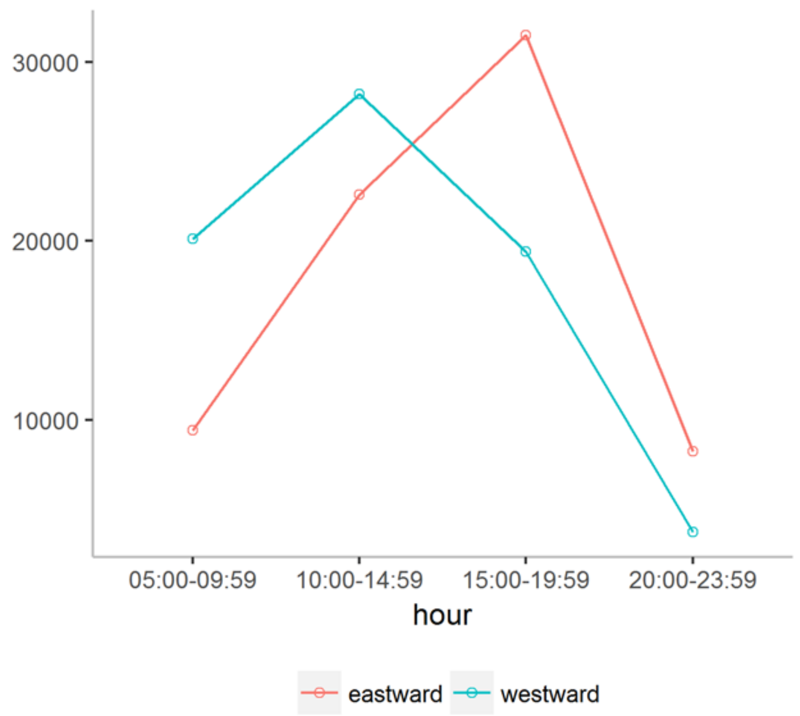
(A) Trip Volume to and from Motijheel area in Sep 2013 (B) Trip Volume Growth to and from Motijheel area from Sep to Nov 2013



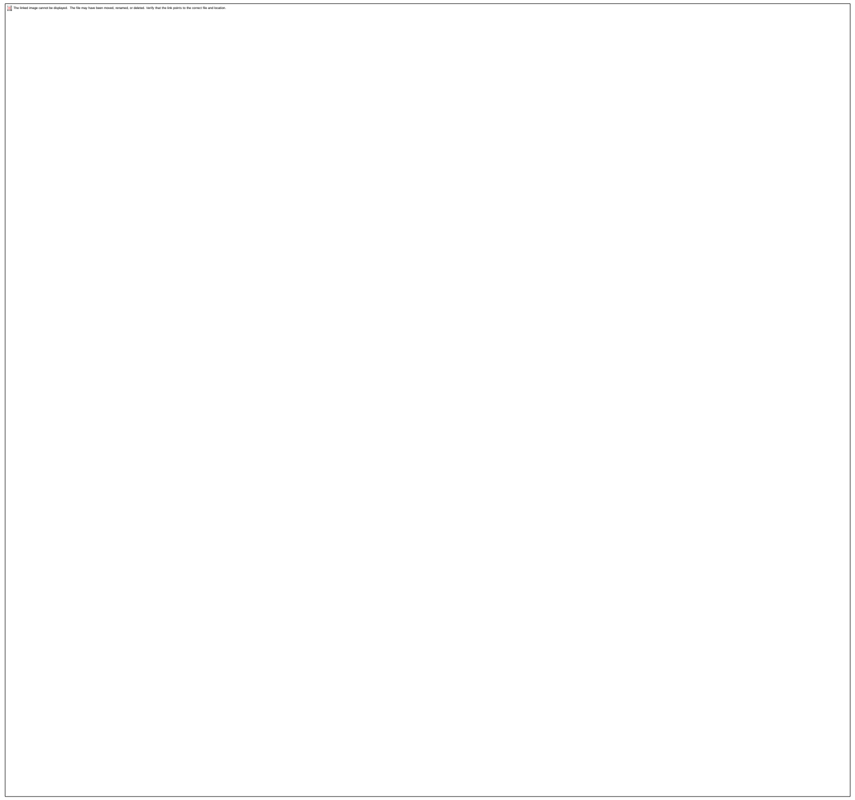
Note: The figures show the heat map of the sum of commuting flows to and from the CBD (Motijheel area; shown in green) in September (Panel (A)) and the growth rate from September to November (Panel (B)). The flyover is marked in black.

Figure 3. Impact of Flyover by Hour and Direction.

(A) Baseline Trip Volume on Flyover Routes



(B) Impact of Flyovers by Direction and Hour



Note: Panel (A) shows the average trip volume for each time of the day and the direction for flyover routes. Panel (B) shows the coefficients and 95% confidence interval of the flyover-impact regression, where flyover dummy is interacted with direction (eastward and westward) and four different time windows (5:00-9:59, 10:00-14:59, 15:00-19:59, 20:00-23:59). For more detail about the specification of the regression, see the footnote of Table 1.

Table 1. Impact of Flyover on Trip Volume and Trip Duration.

Dependent Variable:	Growth Rate of Trip Volume				Growth of Minimum Trip Duration			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flyover	0.238***	0.193***	0.165***	0.257***	-3.965***	-1.773***	-1.082***	-4.458***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.104)	(0.127)	(0.126)	(0.144)
Flyover x Westward				-0.037***				0.975***
				(0.005)				(0.196)
Baseline Volume FE	X	X	X	X	X	X	X	X
Origin FE		X				X		
Destination FE			X				X	
N	1,715,491	1,715,491	1,715,491	1,715,491	625,627	625,627	625,627	625,627
R^2	0.037	0.07	0.07	0.07	0.153	0.181	0.179	0.153

Note: Sample of the regression is defined at the cell tower pair level (origin-destination). In all regressions growth is defined by the change from September to November 2013, before and after the opening of Jatrabari-Gulistan Flyover. Flyover means that the cell tower pair uses at least a part of Jatrabari-Gulistan Flyover identified by the Google Map API. Baseline Volume FE controls for the fixed effects of the number of trips in September divided by 10. *p<0.1; **p<0.05; ***p<0.01.

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

IGC
International
Growth Centre

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UK aid
from the British people

Designed by soapbox.co.uk