Understanding the Impact of Urban Infrastructure: New Insights from Population-Scale Data

Joshua Blumenstock
University of California, Berkeley
School of Information
Berkeley, CA 94720
jblumenstock@berkeley.edu

Danaja Maldeniya
University of Michigan
School of Information
Ann Arbor, MI 48108
dmal@umich.edu

Sriganesh Lokanathan
LIRNEasia
12 Balcombe Place
Colombo, Sri Lanka
sriganesh@lirneasia.net

ABSTRACT

As a growing share of the world's population inhabits cities, a central focus of current development policy has been on building urban infrastructure to support increasing population density. An important component of such policy has been the construction of major roads and highways, which in principle can reduce congestion, promote development of peri-urban areas, and increase labor mobility. However, the actual impacts of such investments have been difficult to evaluate empirically. Here, we use a rich dataset capturing the mobility patterns of roughly 9 million individuals to study the impact of a new super-highway on travel patterns in and around Colombo, the capital of Sri Lanka. Our results indicate that this road had an immediate and pronounced impact on travel patterns: people changed their primary routes of travel, which reduced overall congestion, increased average travel speeds, and reduced the amount of time spent in transit. We further find that the super-highway led to a modest, but statistically significant, increase in the total amount of travel in and around Colombo. We discuss how such insights can inform future policymaking, and point to several promising areas for future research.

CCS CONCEPTS

• Information systems → Spatial-temporal systems; **• Applied computing → Law, social and behavioral sciences**;

KEYWORDS

Transport infrastructure, toll roads, mobile phones, call detail records, ICTD, Sri Lanka

ACM Reference format:
Joshua Blumenstock, Danaja Maldeniya, and Sriganesh Lokanathan. 2017. Understanding the Impact of Urban Infrastructure: New Insights from Population-Scale Data. In *Proceedings of ICTD '17, Lahore, Pakistan, November 16–19, 2017,* 12 pages.
DOI: 10.1145/3136560.3136575

978-1-4503-5277-2/17/11…$15.00
DOI: 10.1145/3136560.3136575

1 INTRODUCTION AND RELATED WORK

Urban areas in developing countries are growing quickly. Roughly one billion of the world's poor live in cities, and by 2030, another 1.4 billion people are expected to be absorbed by the developing world's urban centers [5, 18]. As cities become more densely populated, new challenges are presented around how to manage congestion and overcrowding. A key set of questions in urban development policy thus center around how to build physical infrastructure to support this burgeoning population. As noted in a recent World Bank report, "infrastructure and policy decisions made today will lock cities into urban development patterns for decades to come."[5]

One common policy intervention used to reduce congestion in urban areas is the construction of roads. Globally, the World Resources Institute estimates that global transport investments cost between \$1.4 and \$2.1 trillion annually [28]; in 2014, the World Bank invested \$4.1 billion in urban development [5]. And while a growing body of evidence links rural road construction to improvements in a range of development outcomes including employment [2], school enrollment [22], market access [3], and poverty reduction [21],[1] much less is known about the impact of *urban* infrastructure, particularly in developing countries.

Indeed, there is some ambiguity in the literature as to how new transport arteries might impact congestion. While Baum-Snow [6] shows that highway construction helped draw people from cities in the U.S. to suburban areas, Duranton and Turner [16] find that congestion in U.S. cities is not reduced by new roads, because new roads attract additional drivers, who then clog up the system. In short, many important debates remain unresolved, in part from the lack of reliable empirical evidence. To our knowledge, no previous study has addressed these questions in the context of a developing country.

Part of the challenge in studying the impact of transport infrastructure stems from the historical difficulty of obtaining granular data on urban congestion. Here, we exploit a novel source of data that can shed new light on these longstanding questions. Specifically, we use a large dataset of mobile phone records that capture the trajectories taken by roughly 9 million individuals in Sri Lanka over a period of several months. Importantly, a major new toll highway was opened during the period we observe. This makes it possible to study travel patterns both before and after the opening of the highway, and thus estimate the causal effects of the toll road on traffic and congestion.

A rich body of prior work demonstrates how mobile phone data can be used to model human mobility and migration. Early

[1] Hine et al. [20] provide a review of these and related studies.

work focused on identifying statistical regularities in travel patterns [19, 31, 35], and on quantifying internal migration and population flows [9, 10, 12]. In turn, estimates of mobility derived from phone data have been used to study the propagation of infectious diseases [37, 38], and to study how populations respond to major geopolitical events such as natural disasters and mass gatherings [17, 30, 32]. And while prior researchers have pointed to the potential uses of phone data in urban planning [7, 8, 33], we are unaware of prior work that uses phone data to study the impact of large transportation infrastructure projects.

In the current study, we seek to develop a better empirical understanding of the effect of infrastructure provision on urban mobility. In this sense, our work relates to recent analysis of the role of transport infrastructure in economic growth [4, 15, 36], on the effect of local economic development policies [23, 24], and to a broader literature on economic geography [25]. However, relative to these studies, our goals are much more modest. Our intent is to characterize how a major infrastructure initiative in Sri Lanka impacted the short-run mobility of the region's residents. By providing very fine-grained, dynamic insight into these effects, we hope to lay the foundation for future work that estimates the welfare consequences of these and related urban infrastructure policies.

2 BACKGROUND AND CONTEXT

We begin by briefly introducing the Sri Lankan context and the current technologies used to study traffic and commuting patterins in Colombo. Sri Lanka's agrarian-led economy, and pre-colonial and colonial history, shaped the basic contours of the current transportation infrastructure. The port city of Colombo (population 750,000, Figure 1), which to this day remains the commercial capital of the country, emerged as a multi-modal transport hub as early as the 17th century, when the Dutch constructed internal waterways to transport coconut and spices for export. Under later British rule, as export mainstays shifted towards tea and rubber, railways and roads were built to connect the hill country to Colombo.

After independence in 1948, social welfare policies prioritized rural roads over national roads. As a result, between 1959 and 1990, the rural road network grew from 10,000km to 66,000km, while the national road network grew only from 7,000km to 10,400km [27]. The trade liberalization that started in 1977 also began shifting economic priorities towards a more diversified economy.[2] This in turn shifted the needs of the transport sector. With roads accounting for 94% of the passenger and freight traffic in the country as of 2012, the road network growth has not kept pace with the increased vehicular traffic [14].

With trade liberalization, industrial activity expanded significantly around Sri Lanka's main international airport in Katunayake (32km to the north of Colombo). Katunayake, as well as the adjacent city of Negombo, were connected to Colombo primarily through the A3 road, which we refer to as the "old road" (see Figure 1). As early as 1980, planners and policymakers debated constructing a high-speed link between Katunayake and Colombo, citing economic losses from heavy congestion. Traffic on the old road increased by 200% between 1990 and 2000 [14], when construction

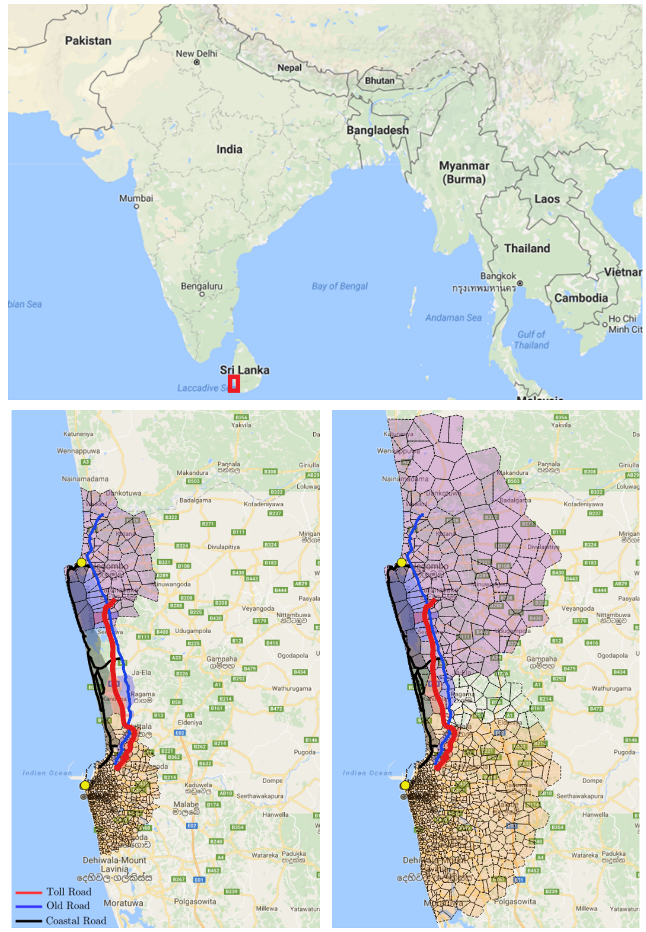


Figure 1: Map of Sri Lankan study area

Notes: Top figure shows location of Sri Lanka, with red box highlighting enlarged regions. Bottom left figure highlights Negombo (north) and Colombo (South) metropolitan areas, with voronoi cells indicated the location of mobile phone towers within 10km of the city centers. Black road is the "coastal road"; blue road is the "old road" (A3), red road is the "toll road" (E3). Right bottom figure shows towers within 20km of city centers.

was started to build a high-speed link to connect Colombo with Katunayke and Negombo. But construction was stopped in 2003 and the contract terminated. In October 2009, construction was restarted and the 25.8km Colombo-Katunayake (E3) Expressway (CKE) was opened to the public in October 2013 – we henceforth refer to this as the "toll road," and it is the impact of this opening that is the focus of this study. The entire project cost nearly USD $321 million, funded mainly through a commercial loan from the Exim Bank of China [14].

Understanding the impact of this new infrastructure is relevant to the broader policy agenda in Sri Lanka. The toll road is the first of a series of planned transport initiatives in the country. In particular, a system of expressways is planned to facilitate travel through and around Colombo, including an Outer Circular (E2) Expressway that is being built in sections. The toll road will also eventually connect to the Colombo-Matara (E1) Expressway and the proposed Colombo-Kandy (E4) Expressway. With current data

[2] As of 2016, the agriculture sector's contributed only 7.5% to GDP, as compared to the industrial sector (27.1%), services (57.%), and taxes (8.4%) [14].

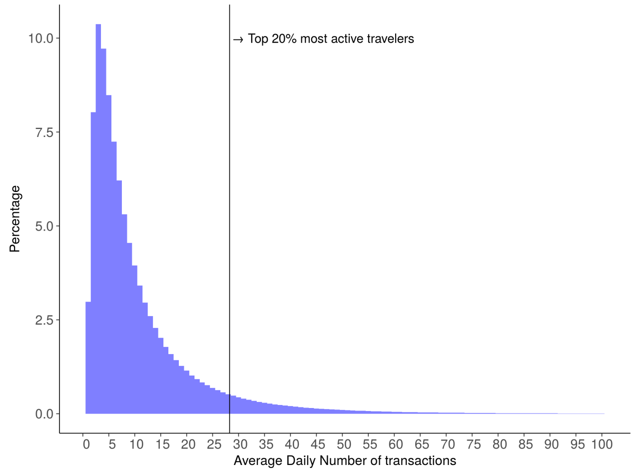


Figure 2: The distribution of subscribers by the average daily number of transactions

and tools, however, policymakers are ill-equipped to evaluate the impact of these major investments.

In particular, traffic congestion is currently measured through a laborious manual process that involves tracking the travel time of specific vehicles between two or more junctions. This process happens infrequently and generally fails to capture the nuances and variation of daily commute patterns. More recently, the Google Traffic API has made data on Colombo available, but uptake of this technology has been low, and our conversations with planners in Colombo reveals a general lack of confidence in, and unawareness of, these data.[3] Traffic speeds are currently captured on major expressways through the use specialized cameras, but this technology is used for enforcement rather than traffic management. There is thus a disconnect between the institutional structures that oversee traffic management (the Road Development Authority) and traffic enforcement (the Police Deartment). This lack of coordination impacts planning efficacy, since automated systems for traffic monitoring fall under the purview of enforcement agencies. Currently, a pilot project is underway in Colombo to develop a coordinated electronic system for a few select key intersections that can provided traffic management as well as traffic enforcement applications. However, such solutions, even if successful, would require significant investment in physical devices and human resources to manage deployment and maintenance.

3 DATA

Our empirical results are based on the analysis of travel patterns in and around Colombo, before and after the opening of the toll road. To measure these travel patterns, we build on a growing literature that uses mobile phone metadata to model human mobility [9, 12, 19, 35]. Specifically, we use a large database of pseudonymized Call Detail Records (CDRs) covering approximately 9 million individuals from an unnamed operator in Sri Lanka. The dataset contains 4 contiguous months of CDRs, from August 2013 to November 2013 inclusive, and thus overlaps conveniently with the October 2013 opening of the toll road. This dataset is a subset of a larger dataset that extended up to November 2013 and was not specifically collected for this analysis. As a result our analysis is limited to exploring the effect of the toll road within the relatively short time period of up to 1 month of its opening.

The CDRs contain the metadata passively collected by the mobile network whenever a subscriber uses the mobile phone to make or receive a phone call, send or receive a text, or when initiating a data session. For each such event, we observe the following bits of information:

(1) A unique identifier for the calling/sending party.
(2) The date and time at which the event was initiated.
(3) The ID of the cellular antenna the subscriber was connected to at the time of the call. Each antenna ID is mounted on a mobile phone tower, which we can dereference to a physical (latitude, longitude) location.

Other information contained in the CDR is discarded. Our analysis focuses only on call and data events, as the text message CDR was not made available by the operator. Unique identifiers do not contain personally identifying information, as they were pseudonymized by the operator (i.e., each phone number in a CDR was replaced a unique computer-generated identifier).

Figure 2 shows the distribution of unique events observed per day for each of the subscribers in our dataset. The average subscriber is involved in 12 calls or data requests per day, but this distribution is heavily skewed (median = 7, SD = 30). In later calibration tests, we will restrict analysis to the subset of subscribers who are most active on the network, as it is easier to minutely trace their mobility patterns - these individuals are defined as those involved in 28 or more events per day (i.e., the top 20% of the activity distribution).

Following standard practice in the literature, we use the GPS coordinates of the mobile phone towers to construct a voronoi tessellation of the physical landscape. This divides the landscape into regions of approximate tower coverage, as can be seen in the bottom subfigures of Figure 1. As summarized in Table 1, there were approximately 4,000 towers owned by this operator in Sri Lanka in 2013. Within 10km of the center of the two metropolitan regions of Colombo and Negombo, there were 800 and 100 towers, respectively, with tower density considerably higher in the capital city of Colombo.

4 METHODS

To understand the impact of the toll road, we first need a method to quantify travel patterns from the mobile phone data described above. In particular, we wish to estimate the number of people traveling between the two zones connected by the road and the specific route chosen by each of those travelers. To understand the impact on congestion, we also want to calculate the total duration of each trip, as well as the velocity of travel. Here, we describe the methods used to (1) extract "trips" from the CDR, corresponding to travel by subscribers between the two regions; (2) Assign trips to routes, i.e., to determine which of the three routes connecting Colombo and Negombo was used for travel; (3) Measure trip duration and travel

[3]Information gathered by S. Lokanathan in conversations with the Western Region Metropolis Authority and faculty in the Department of Transport and Logistics Management at the University of Moratuwa.

Region	# Towers	Avg. Area (km^2)	Area SD
Sri Lanka	≈4000	18.19	36.07
Colombo (10km)	≈800	0.17	0.25
Negombo (10km)	≈100	2.1	1.63
Colombo (20km)	≈1000	0.42	0.67
Negombo (20km)	≈200	3.3	2.64

Notes: Each row indicates the number of towers in a given geographic region, as well as the average size of the voronoi cell and the standard deviation of that size. Row 1 includes all of Sri Lanka; rows 2-3 indicate the towers within 10km of the city centers; rows 4-5 indicate towers within 20km of the city centers. Approximate numbers shown (first column) at the request of the operator.

Table 1: Characteristics of voronoi catchment areas

speed; and (4) estimate the effect of the opening of the toll road on number of trips taken, trip duration, travel speed.

4.1 Trip Extraction

Our current focus is on understanding the impact of the toll road on the trips taken between Negombo and Colombo (we discuss in Section 6 the potential for scaling this approach to travel indirectly impacted by the road). We begin by demarcating the regions of Colombo and Negombo as the area within 10km of each respective citiy center, and later show that our qualitative results do not change if we use a radius of 20km.[4]

A trip between Colombo and Negombo is then defined as a sequence of multiple CDR events from a single user, where one event is associated with a tower in the Colombo region, and another event is associated with an event in the Negombo region; a trip from Negombo to Colombo is defined analogously. We focus on trips that occur within a single day. Issues of measurement error and bias are discussed separately below.

Formally, let $\mathbf{s}$ be a subscriber in the dataset and C and N be the set of mobile towers in the regions of Colombo and Negombo, respectively.[5] Denote by $\mathbf{P}$ the chronologically ordered sequence of locations observed in the CDR of $\mathbf{s}$, i.e., $\mathbf{P} = \{p_i\}_0^n$. We consider $\mathbf{s}$ to have made a trip between the two regions if $\mathbf{s}$ is observed at location l_1 at time t_1 and at location l_2 at time $t_2 (> t_1)$ where $l_1 \in C(N)$ and $l_2 \in N(C)$, such that there are no other observations in N or C at a time t where $t_1 < t < t_2$.

In other words, a trip $\mathbf{t}$ satisfies $p_0 = l_1$ and $p_n = l_2$, and can be represented as $\mathbf{t} = \{\mathbf{s}, l_1, t_1, l_2, t_2, \mathbf{P}\}$, where

- $\mathbf{s}$ = Subscriber
- l_1 = Trip origin
- l_2 = Trip destination
- t_1 = Trip start date and time
- t_2 = Trip end date and time

Using this definition of a trip, we identify a total of 9,181,894 trips between Colombo and Negombo in the 4-month span of data.

[4]The 10km radius is preferred as it corresponds to what local citizens roughly consider to be the metro regions of the two cities, see Figure 1. In results available on request, we also test the robustness of later results to radii of 5km and 15km.

[5]The city centers for Colombo and Negombo ware defined as Colombo = (lat - 6.9344, lon - 79.8428), Negombo = (lat - 7.2111, lon - 79.8386)

4.2 Route Assignment

Having established the set of relevant trips between Colombo and Negombo, we wish to determine which of the three major routes the subscriber was most likely to have traveled (see Figure 1). Intuitively, the idea is to look at sequence of locations at which all $\mathbf{s}$ was sighted between the start (p_0) and end (p_n) of the trip, and determine which route the majority of those locations are nearest to. In practice, developing this algorithm is not trivial, since the different routes traverse a different number of voronoi cells,[6] the average size of the cell is different along different routes, and often multiple routes traverse the same voronoi cell. And most problematically, the set of sightings for any given trip is often quite sparse. Thus, there exist a variety of reasonable approaches one might take to inferring route choice from the sequence of CDR events related to the trip.

We therefore develop four related algorithms for route assignment. The most conservative *mode heuristic* considers only those towers that are unique to a single route, and assigns a trip to the route that contains the majority of unique towers are visited. The related *percentage heuristic* accounts for the fact each routes has a different number of unique towers, and and assigns a trip to the route that contains the highest fraction of possible unique towers. Finally, for both the mode and percentage heuristic, we also develop an aggressive *Bayesian* method that helps classify trips with few intermediate sightings. Intuitively, the Bayesian method allows non-unique towers to assign probability weights to routes when those non-unique towers are frequently observed on other trips that occur with high probability on specific routes. Formal definitions follow.

Mode Heuristic. For a trip $\mathbf{t}$, let the number of unique mobile phone towers observed in $\mathbf{P}$ for the toll road, old road and the coastal road be n_{toll}, n_{old}, and $n_{coastal}$. Then the road $R(t)$ assigned to $\mathbf{t}$ is simply

$$R^{mode}(t) = \begin{cases} \arg\max_i n_i, & \text{if } \arg\max_i n_i \text{ is unique} \\ \text{NA}, & \text{otherwise} \end{cases} \tag{1}$$

Where $i \in \{toll, old, coastal\}$. This approach classifies 3,722,721 (40.54%) of all trips; the others are unclassified.

Percentage Heuristic. Given that each route traverses a different number of unique mobile phone towers, the mode heuristic is likely to biased towards roads that have more unique towers. The percentage heuristic approach accounts for the variation in the number of unique towers among the three roads by considering the observed number of unique towers for each road relative to the total number of unique towers available to that road. Therefore, a trip $\mathbf{t}$ is assigned a road $R(t)$, as follows.

$$R^{pct}(t) = \begin{cases} \arg\max_i \frac{n_i}{T_i}, & \text{if } \arg\max_i \frac{n_i}{T_i} \text{ is unique} \\ \text{NA}, & \text{otherwise} \end{cases} \tag{2}$$

Where, $i \in \{toll, old, coastal\}$ and T_i is the number of towers unique to road i. This heuristic classifies 4,105,624 (44.71%) of all identified trips.

[6]Based on the base station coverage regions shown in Figure 1, there were 15,19 and 32 mobile phone towers that served only parts of the toll road, old road and the coastal road respectively.

	Mode		Percentage		Mode + Bayes.		Percentage + Bayes.	
Panel A: Prior to toll road opening								
Old road	7,061	(922)	7,545	(1,000)	19,944	(2,089)	18,871	(1,942)
Coastal road	3,275	(632)	3,249	(625)	4,693	(797)	4,551	(774)
Toll road	1,449	(219)	2,318	(421)	2,047	(269)	3,262	(473)
Unclassified	14,899	(1433)	13,572	(1,242)	0	(0)	0	(0)
Total	26,684	(3,097)	26,684	(3,097)	26,684	(3,097)	26,684	(3,097)
Panel B: After toll road opening								
Old road	5,512	(741)	5,894	(789)	16,523	(1,892)	15,038	(1,711)
Coastal road	2978	(497)	2,948	(490)	4,576	(705)	4,425	(686)
Toll road	3,016	(472)	3,895	(605)	7,248	(956)	8,884	(1,152)
Unclassified	16,841	(1,852)	15,610	(1,688)	0	(0)	0	(0)
Total	28,347	(3,472)	28,347	(3,472)	28,347	(3,472)	28,347	(3,472)

Notes: Pairs of columns indicate the average (and standard deviation, in parentheses) number of trips per day on each route, across all days, separately for the period before and after the road opening. Each pair corresponds to a different route assignment algorithm, as described in Section 4.2

Table 2: Trips per day, according to different route classification algorithms

Bayesian (Probabilistic) Route Assignment. The approaches described above are only able to assign routes to a fraction of all trips observed in the dataset. We thus develop a probabilistic method that leverages trips previously classified with the heuristic approach to bootstrap the classification of the remaining trips.

(1) We consider the road assignments by the heuristic step as a soft assignment for the purpose of this step. Therefore, the weight assigned to a trip **t** with respect to a road i, $W(i|t)$, is estimated as follows.

$$\texttt{Mode:} \quad W(i|t) = \frac{n_i}{\sum_i n_i} \tag{3}$$

$$\texttt{Percentage:} \quad W(i|t) = \frac{\frac{n_i}{T_i}}{\sum_i \frac{n_i}{T_i}} \tag{4}$$

(2) The weights are used to estimate the expected number of already labeled trips and the prior probability of a trip being on each road before and after the opening of the toll road. Let $E_p(i)$ be the expected number of trips for a road $i \in \{toll, old, coastal\}$, for the period $p \in \{before, after\}$ and $P_p(i)$ be the corresponding prior probability. Let $T_{p,heuristic}$ be the total number of trips during p that were labeled by the heuristic method.

$$E_p(i) = \sum_t W(i|t) \tag{5}$$

$$P_p(i) = \frac{E_p(i)}{T_{p,heuristic}} \tag{6}$$

(3) We estimate the conditional probability of a non-exclusive mobile phone tower, b_j, being observed during a trip **t** on road i for each non-exclusive mobile phone tower and road pair. These probabilities are estimated only using trips after the toll road opened that were classified by the heuristic method.

$$P(b_j|i) = \sum_t \begin{cases} W(i|t), & \text{if } b_j \text{ observed in } t \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

(4) Finally, we classify each trip t that was not classified by the heuristic method using the following Naive Bayesian classification formula

$$R^{Bayes}(t) = \arg\max_i P_p(i) \prod_j P(b_j|i) \tag{8}$$

The Bayesian variants of the heuristics were used to classify trips that remained unclassified after applying the original heuristics. The Bayesian variant of the mode heuristic classifies 5,459,173 (59.45%) of all trips. The Bayesian variant of the percent heuristic classifies 5,076,270 (55.29%) of all trips.

Complete details on the assignments made by the different route assignment algorithms are shown in Table 2. In general, the Bayesian approaches are able to classify a larger number of trips; but show an empirical bias toward classifying trips along the Old Road. Since this estimate will, if anything, lead us to conservatively underestimate the impacts of the toll road opening, we use this as the specification to calculate the primary results in Section 5. In separate robustness checks in Section 5.3, we show that, as expected, the impact of the toll road looks even larger if we use one of the alternative route assignment algorithms.

4.3 Trip characteristics

For each classified trip we calculate several characteristics:

(1) *Number of Trips*: The number of trips for a day d on a road i is the number of trips on i whose start date time t_1 is during d (Table 2).
(2) *Number of Travelers*: The number of travelers for a day d on a road i is the number of unique subscribers traveling on each day.

(3) *Travel Time*: We measure the travel time of each trip as the difference $t_2 - t_1$ between the start and end of the trip. The daily travel time for road i is then estimated as the median value of the travel times assigned to i on each day of the observation period. We note that this is almost certainly an overestimate of the actual travel time, since people may make phone calls long before departing or after arriving. However, since our interest is on understanding relative changes in travel time, we make the assumption that the systematic error in travel time is comparable for each of the three routes.

(4) *Travel Velocity*: To estimate the travel velocity for each trip, we compute the "as the crow flies" physical distance $D(l_1, l_2)$ between l_1 and l_2, and calculate

$$\text{Velocity} = \frac{D(l_1, l_2)}{t_2 - t_1} \tag{9}$$

As a result of using of the "as the crow flies" distance our velocity measurements underestimate actual travel velocity of trips. The daily travel velocity for road i is then estimated as the median value of the travel velocities assigned to i on each day of the observation period.

Intra-day variation. Later, we will also disaggregate our daily results by time of day, to better understand if there are certain periods when travel is most impacted by the toll road. For this analysis, we estimate median travel time for trips initiated during each 20 minute interval of the day, on each route, both before and after the toll road opening. In that analysis, we will compare the two-week period one month before the toll opening to the two-week period one month after the toll road opening. To increase the precision of these estimates, we employ a recent technique to derive inter-city travel times with CDR [26]. This method better accounts for extreme outliers in travel times estimated from CDR, which are less likely to impact our daily median estimates, but become more important in the intra-day estimates where data becomes more sparse.

4.4 Estimation

Finally, to more precisely quantify the effect of the toll road opening, we use a fixed effect model of the form:

$$V_{t,d} = \beta(POST_t) + \delta t + \mu_d + \epsilon_{t,d} \tag{10}$$

where $V_{t,d}$ is the dependent variable of interest (i.e., number of trips, number of users, travel time, or travel velocity) on a given road (i.e., we run separate regressions for each of the three roads), and $POST_t$ is an indicator variable that takes the value one for days t that are on or after the opening and zero otherwise. The variable t controls for route-specific time trends, for instance if traffic is stochastically increasing or slowing over time, and μ_d is a day of week fixed effect that accounts for the fact that traffic patterns may be different on different days of the week. We are most interested in the coefficient β, which indicates the route-specific impact of the toll road opening, after controlling for all of the factors discussed above.

5 RESULTS: IMPACT OF THE HIGHWAY

We turn now to analyzing the impact of the opening of the E3 super-highway on travel and mobility patterns in Sri Lanka. We begin by examining how the highway impacted the transit routes chosen by individuals traveling between the two affected cities of Colombo and Negombo, and then discuss how it affected the speed of travel and total amount of time spent in transit. Finally, we present several robustness checks to "stress test" the main results and the modeling assumptions made along the way. Quantitative results are summarized in Table 3, and explained in greater detail below.

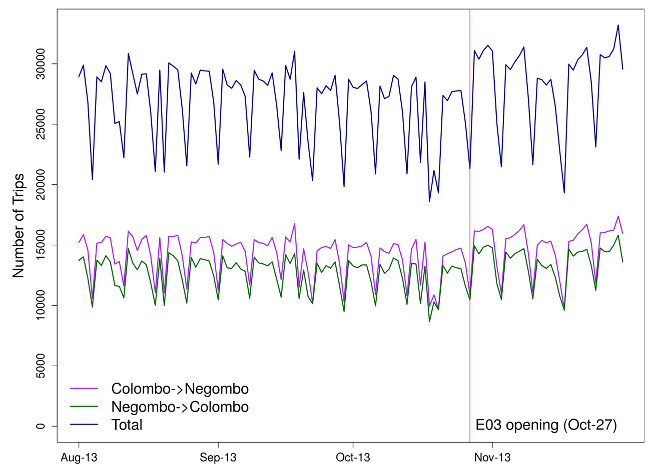


Figure 3: Total trips taken between Colombo and Negombo

5.1 Travelers, Trips, and Route Choice

Our first empirical result is to observe that the total number of trips taken between the two cities modestly increased following the opening of the toll road. This can be seen in Figure 3, which shows a time series of the total number of trips across all routes over time (dark blue line). The weekly cyclicality is evident, with a regular lull in travel on Sundays. The red vertical line marks the opening of the toll road, and the small increase in travel following the opening is evident.

The aggregated data in Table 2 indicates an increase in the post-opening period of roughly 5.8%, from 26,684 to 28,347 observed trips per day. We further use model (10) to control for day-of-week effects and stochastic time trends; results are presented in the first column of Table 3. These estimates indicate that an additional 3,052 trips were made after opening, which is comprised of 5,255 additional trips on the toll road, 2,023 fewer trips on the old road, and 150 fewer trips on the coastal road (though this latter effect is not statistically significant). Similarly, the second column of Table 3 indicates that 2,981 new unique travelers were observed per day, in the period after the toll road opened.

The pronounced shift in the routes taken by travelers can also be seen in Figures 4 and 5, which show the number of unique trips and travelers between Colombo and Negombo on each of the three main roads. Travel on the toll road sharply increases (blue line), and is offset by a corresponding decrease in travel on the old road (black line). There is thus clear evidence of substitution by certain travelers onto the faster (and more expensive) toll road. However, travel on the third coastal road is largely unaffected. These visual trends are

	Number of trips	Number of travelers	Median travel time	Median velocity
Toll Road	5,225***	4,423***	-25.62***	8.53***
	(145)	(123)	(1.36)	(0.39)
Old Road	-2,023***	-1,417***	-7.57***	1.82***
	(390)	(310)	(0.71)	(0.27)
Coastal Road	-150	25	0.09	0.42
	(134)	(107)	(0.67)	(0.11)

Notes: Each cell represents the coefficient β from a separate fixed-effects regression of a traffic outcome (column heading) for a given road (row heading) on an indicator variable of the period post-toll road opening, after controlling for time trends and fixed effects (see model 10). All coefficients other than β are omitted for clarity. Standard errors shown in parentheses. *** p < .001, ** p < .01, * p < .05

Table 3: Quantifying the impact of the toll road opening on traffic patterns on each road

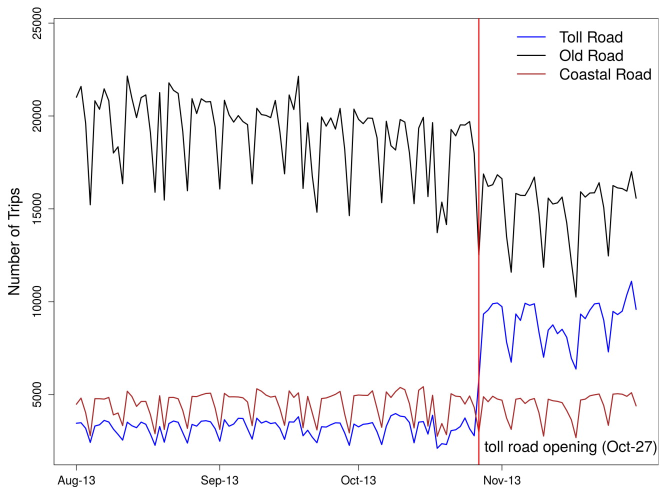


Figure 4: Daily trips taken on each route, over time

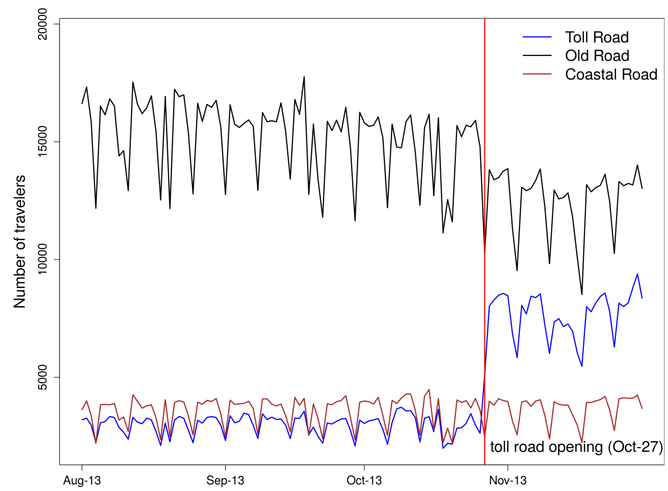


Figure 5: Number of unique travelers on each route

supported by the quantitative evidence in Table 3, which indicate that the decline in travel on the old road explains approximately 40% of the observed increase in travel on the toll road after opening. The remaining 60% of travel on the toll road appears to be driven by the net increase in travel discussed earlier.

It is worth noting the non-trivial number of trips assigned to the toll road prior to the toll road's opening. This is evidence of classification error by the route assignment algorithm, and can help us better understand possible measurement error in the figures and tables we present. In practice, these trips most likely reflect travel by subscribers who reside in areas covered by mobile phone towers that also uniquely cover the toll road once it opens. Given that these subscribers are likely to continue to travel between the two regions after the opening, the change in travel volume assigned to the toll road since the opening is likely to be a better estimate and at worst an underestimate of the actual volume of travel on the toll road.

5.2 Travel Time and Speed

The opening of the toll road also significantly reduced congestion between Colombo and Negombo. Figure 6 shows the average (median) travel velocities on each of the three roads, over time. While we caution against interpreting the absolute value of the calculated velocity, since this is subject to considerable measurement error due to the potential delay between when people start and end their trip and when they make a phone call from their origin and destination (see Section 4), the relative change in speed of travel is pronounced. First, travel speed on the toll road is roughly 30% faster than on all other roads, and roughly 50% faster than travel on the other roads prior to the toll road's opening. Second, travel speeds on the alternative routes (and in particular the old road) also increased after the opening. We attribute this decrease to the reduced congestion caused by a reduced load on the alternate routes.

The increase in travel velocity equates to a decrease in the amount of time spent traveling, as shown in Figure 7. While the absolute travel time is likely an overestimate for the same reasons articulated earlier, the relative shift is again quite striking. If we momentarily suspend disbelief and take these estimates literally, with roughly 9,000 trips on the toll road each being shortened by roughly 20 minutes, this indicates a daily reduction of 3,000 commuting person-hours. This is likely a dramatic underestimate, since it does not account for travelers who do not own a phone, who use a phone on a different network, or who simply do not use their phones immediately before and after traveling. It also does not account for reduced congestion on the network of roads impacted

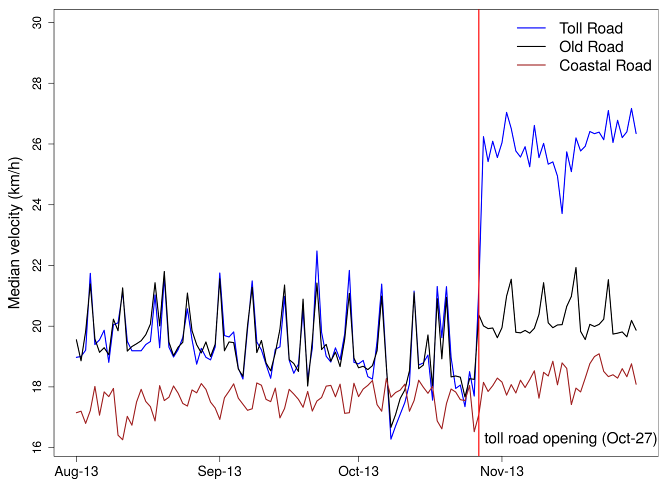


Figure 6: Median daily travel speed on each route

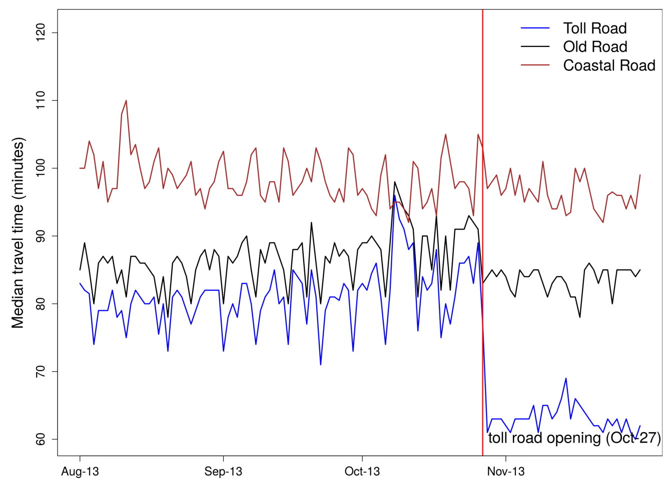
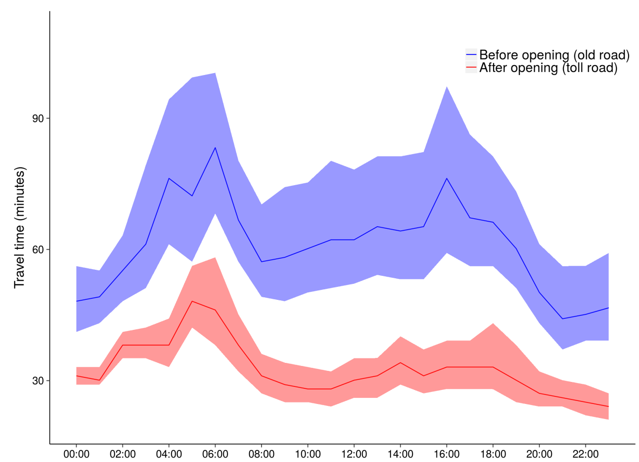


Figure 7: Median daily trip duration on each route

by this new artery. We return to these and related issues in the discussion in Section 6 below.

Finally, Figure 8 disaggregates the impact of the toll road on the trip length by time of day. Here, we show the median travel duration for travel between Colombo and Negombo prior to the toll road opening (blue line) as well as the median travel duration after the opening (red line). Trips are assigned to 1-hour intervals based on the time of day at which the trip began. We observe that while travel times after the opening are shorter at all times of day, the effect is most pronounced at times after the morning commute hour. Qualitatively, there are still traffic jams on the toll road in the morning; these jams are not as bad as the jams on the old road, but clearly the volume of traffic still exceeds the maximum capacity of the road. By contrast, outside of this morning commute period - and even during the evening commute - travel time on the toll road appears to consistently occur at the minimum duration possible.



Notes: Shaded areas denote 5^{th} and 95^{th} percent (bootstapped) confidence intervals

Figure 8: Impact on trip duration, by time of day

5.3 Robustness and Calibration

Since one contribution of this paper is to present a new method for using mobile phone data to study the impact of new transport infrastructure on travel behaviors, we believe it is important to understand how several key modeling decisions have affected the main empirical results. In particular, we investigate the importance of the method used to classify trip routes from the original call detail records (subsection 5.3.1), the potential bias from data sparsity (subsection 5.3.2), and the importance of how we define origin and destination regions (subsection 5.3.3). The sum total of these tests, as well as others omitted due to space constraints,[7] lead us to conclude that the main results reported in this paper – in particular the significant changes in route choice and large reduction in travel time – are robust to a variety of plausible modeling assumptions. However, the exact magnitude of these changes estimated from the data – an exercise that we leave for future work and discuss below – will indeed depend on several modeling decisions. Other potential sources of bias and related concerns that cannot be directly tested with our data are discussed in Section 6.

5.3.1 Route Assignment Algorithm. Figure 9 (left) shows the daily travel time estimates for the three alternative roads based on the "percentage" route assignment algorithm describe in Section 4.2. We note that the median travel time estimated from this algorithm is slightly lower across the board, presumably due to the fact that we are able to classify fewer trips than with the Bayesian algorithm. However, the qualitative picture of the impact of the toll road are unchanged. In particular, travel time on the toll road dropped precipitously after the road's opening, while travel time on the other two roads showed only modest reductions.

[7]In results available on request, we test: (i) alternative route classification algorithms; (ii) several different thresholds for determining the regions denoted by Colombo and Negombo; (iii) disaggregating effects by weekday and weekend; and (iv) different methods for removing outliers in the data.

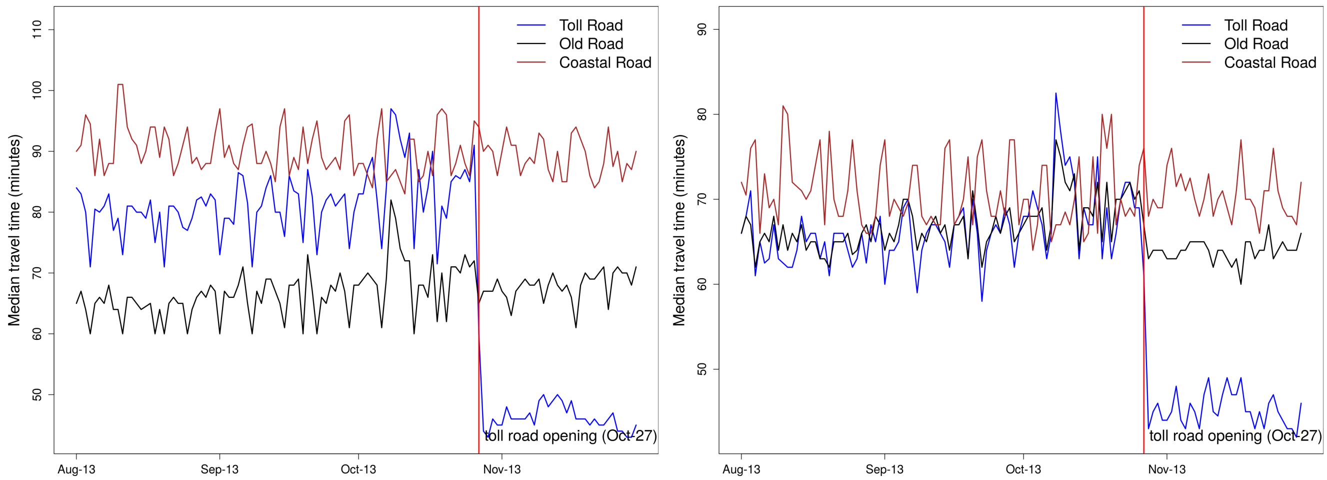


Figure 9: Testing the robustness of main results. Left figure uses a mode heuristic route assignment algorithm. Right figure uses the population of high activity users. Both figures are most directly comparable to Figure 7

5.3.2 High-Frequency Users. One concern with these results is that we are inferring travel times from individuals who, on average, are only sighted on the network 12 times per day (see Section 3). Since our estimates rely on the traveler using his or her phone before and after the trip, infrequent mobile phone users will appear to take very long trips in our data. If this effect is evenly distributed across routes, it should not impact our ability to infer changes in the relative trip length between routes, but it would impact our estimate of the absolute trip length and travel velocity.

To address this concern, we restrict our analysis to the set of subscribers who are observed frequently in the data, and for whom our estimates are likely to be more accurate. Empirically, we take the top 20% of active subscribers, which is equivalent to those individuals involved in at least 28 events per day (see Figure 2). Figure 9 (right) shows daily travel time estimates using the top 20% most active subscribers. Two points are most salient. First, estimates of total travel time are consistently lower on all routes at all times, reflecting the intuition that we are more likely to observe high-activity subscribers immediately before and immediately after the trip. Second, despite these differences, the key qualitative point – that the opening led to a reduction in travel time of roughly 15% on the toll road – remains unchaged.

5.3.3 Region of Analysis. Finally, we consider whether the observed effects are impacted when considering travel over a more extensive region beyond the city centers. We estimated the travel characteristics between Colombo and Negombo with the city regions being bounded by the coverage of mobile phone towers within 20km from the city centers. Since the geodesic distance between the two city centers is less than 40km we introduced a small neutral region between the two cities as seen in Figure 1 (bottom right). Figure 10 shows the daily travel times for the three alternative roads based on these larger regions. We observe that the effect of opening the toll road is similar but smaller compared to our main set of results. This decline in the effect is intuitive as the advantage offered by the toll road is less significant for travel originating and ending further away from it.

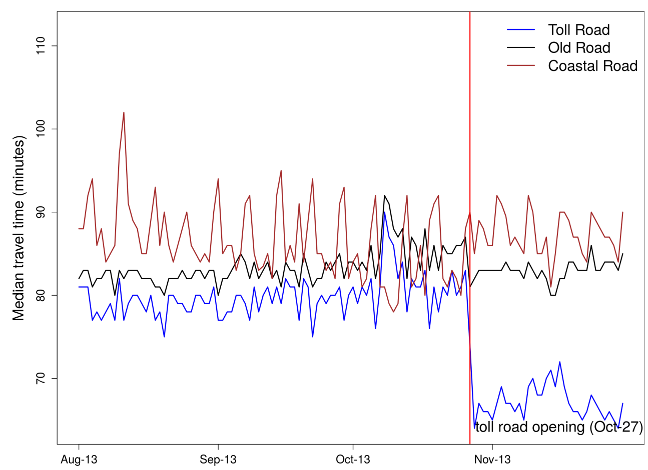


Figure 10: Robustness to 20km radius

6 DISCUSSION AND LIMITATIONS

The results thus far indicate that the opening of the E3 superhighway had an immediate and pronounced impact on travel patterns near Colombo. We have also shown that this qualitative finding does not change under a variety of alternative modeling approaches. Here, we address several additional factors that we believe are important in interpreting these results, including possible sources of bias and important limitations to consider in future work.

Selection Bias and Representativity. The inferences we make about travel patterns in Sri Lanka are based on data produced when mobile phone subscribers use their phones. This leads to three distinct sources of potential bias:

(1) *Phone Ownership:* In Sri Lanka and other developing economies, mobile phone owners are not a representative sample of the at large population. Rather, phone owners tend to be wealthier, better educated, and are more likely to be male [11].
(2) *Mobile Phone Operator choice:* Our data comes from a single mobile phone operator that is one of the largest in Sri Lanka; thus we do not observe the communication or travel patterns of those that use other networks. When compared to the census population we find that our dataset on average covers roughly 55% of the population in the regions considered in this analyses.
(3) *Phone Usage:* We observe peoples' locations only when they are active on the mobile phone network. In general, this causes us to underestimate the mobility of people, and overestimate the duration of the average trip (see also the discussion in Section 5.3.2). More insidiously, to the extent that peoples' phone use changes systematically while traveling on specific roads, such changes would bias our estimates of the impact of the new expressway. For instance, some people may need to focus more when driving on expressways and would therefore be less likely to use their phones in transit. Alternatively, some people may see the comfort of the expressway as as an ideal time to catch up on phone calls.

While prior work has shown that certain measures of mobility derived from mobile phone data approximately generalize to the entire population [9, 29, 39], the specific points highlighted above are important limitations that have not been sufficiently addressed. In our context, we are seeking third-party sources of data on traffic in Sri Lanka to calibrate our phone-based estimates with traditional sources of "ground truth" data.

Route Classification. One of the most important and challenging methodological components of this work was determining how to assign individual travelers to physical routes. Section 4.2 discusses several plausible solutions, and Section 5.3.1 shows the general robustness to this design decision. In the end, however, our choice of the algorithm was largely arbitrary: we chose the algorithm that best reflected our intuition regarding how route assignment from call detail records, and was rooted in prior work on mobility and in Sri Lanka [26, 27]. We hope that future work can tackle this question in a more principled and systematic way, ideally through validation with a separate source of trusted data. For instance, if survey data could indicate total or relative flows on each route on specific days, such data could be used to help choose among the various route classification algorithms, or better yet, to train a non-parametric route classification algorithm.[8]

Short- vs. Long-Run Impacts. While our current focus is on the immediate impact of the expressway's opening, economic theory suggests that over time, the increased convenience of travel will induce more people to take to the roads, eventually leading to steady-state levels of congestion similar to the prior equilibrium [16]. Here we are constrained by the fact that we have only two months of data after the road's opening. In that period, we see no evidence of the effects attentuating, but note that in order to address these more fundamental economic questions would require a significantly longer period of observation (and ideally, multiple new transport infrastructures to evaluate).

Local vs. Systemic Effects. As shown in Figure 1, our analysis is based on changes in the travel patterns of people moving between Colombo and Negombo, the two urban regions most impacted by the new road. In this sense, we are estimating a local average treatment effect, and ignoring more systemic, general equilibrium effects. On the one hand, this likely leads us to overestimate the impact of the road on each individual traveler, since estimates of quantities like median velocity and trip time are based on the most impacted travelers. On the other, we are likely underestimating the aggregate effects to the broader Sri Lankan society, since reducing congestion along certain key arteries likely has positive externalities to the larger transport network.

Privacy Concerns. Our study relies on using the pseudonymized mobile phone usage logs of millions of subscribers, who never explicitly consented to have their data used in this study. While we went to great lengths to ensure that our analysis minimized the potential privacy risks to human subjects,[9] prior work has shown that even pseudonymized data such as that used in our study can uniquely identify individuals [13]. While a discussion of the ethical considerations involved in such research is well beyond the scope of this paper, our determination was that the potential insights gained from this line of inquiry justified the minimal risk posed to mobile subscribers.

7 CONCLUSION

In studying the opening of the E03 super-highway in Sri Lanka, this paper makes three contributions. The first is methodological, by demonstrating the potential for pseudonymized mobile phone data to provide new insight into the impact of urban infrastructure on travel behaviors in developing countries. We develop and calibrate a set of methods for transforming mobile phone call detail records into practical measurements that can aid in policy evaluation. Related, this paper makes an empirical contribution, by showing how these methods can reveal detailed information about the short-term impacts of new transport infrastructure on urban congestion, and specifically on the routes chosen by travelers and the speed of travel across directly impacted roads. In the context of Sri Lanka, to our knowledge this represents the first and only piece of empirical evidence on the effects and effectiveness of the $321 million investment in the E03.

The final contribution, and perhaps the most important one, is to lay the groundwork for what we see as an important research agenda around understanding the *welfare impacts* of urban infrastructure. Given a longer panel of phone data, we believe the most important questions focus on quantifying the short- and long-term

[8] Additional subtleties arise from the non-random location of the mobile phone towers, which are used to construct the Voronoi division of the physical space that in turn forms the basis for route classification. Our "percentage" heuristic (see Section 4.2) helps address this concern, but does not account for the endogenous response of physical infrastructure to the presence of additional subscribers. If the physical network co-evolves with human congestion, this will require a more nuanced, dynamic approach to route classification.

[9] Among other measures: all personally identifying information was stripped from the data prior to analysis; all analysis was performed on a secure server in Sri Lanka, with no data ever leaving the country; all results presented involve aggregates over a large number of subscribers; IRB approval was obtained (including a waiver for informed consent) from the U.C. Berkeley Committee for the Protection of Human Subjects.

economic costs and economic benefits, as well as the distributional implications (i.e., how different subgroups are differentially affected), of these massive investments. Related work provides a framework for modeling these welfare impacts [1, 15, 34], but has not benefited from the fine-grained measurements that we develop in this paper. Combining the deep theoretical foundations of those models with the rich data and methods now at our disposal has the potential to contribute new insight into pressing questions that affect the lives of hundreds of millions of individuals in urban developing contexts.

REFERENCES

[1] Treb Allen and Costas Arkolakis. 2014. Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics* 129, 3 (Aug. 2014), 1085–1140. DOI:https://doi.org/10.1093/qje/qju016
[2] Sam Asher and Paul Novosad. 2014. *The Employment Effects of Road Construction in Rural India*. unpublished. http://www.isid.ac.in/~epu/acegd2014/papers/SamAsher.pdf
[3] Sam Asher and Paul Novosad. 2016. Market access and structural transformation: Evidence from rural roads in India. *Working Paper* (2016).
[4] Abhijit Banerjee, Esther Duflo, and Nancy Qian. 2012. *On the Road: Access to Transportation Infrastructure and Economic Growth in China*. Working Paper 17897. National Bureau of Economic Research. http://www.nber.org/papers/w17897 DOI: 10.3386/w17897.
[5] World Bank. 2014. Urban Development: Sector Results Profile. (2014). http://www.worldbank.org/en/results/2013/04/14/urban-development-results-profile
[6] Nathaniel Baum-Snow. 2007. Did Highways Cause Suburbanization? *The Quarterly Journal of Economics* 122, 2 (May 2007), 775–805. DOI:https://doi.org/10.1162/qjec.122.2.775
[7] Richard Becker, RamÃşn CÃąceres, Karrie Hanson, Sibren Isaacman, Ji Meng Loh, Margaret Martonosi, James Rowland, Simon Urbanek, Alexander Varshavsky, and Chris Volinsky. 2013. Human Mobility Characterization from Cellular Network Data. *Commun. ACM* 56, 1 (Jan. 2013), 74–82. DOI:https://doi.org/10.1145/2398356.2398375
[8] R. A. Becker, R. Caceres, K. Hanson, J. M. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky. 2011. A Tale of One City: Using Cellular Network Data for Urban Planning. *IEEE Pervasive Computing* 10, 4 (April 2011), 18–26. DOI:https://doi.org/10.1109/MPRV.2011.44
[9] Joshua Evan Blumenstock. 2012. Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda. *Information Technology for Development* 18, 2 (2012), 107–125.
[10] Joshua E. Blumenstock, Ramkumar Chokkalingam, Vijay Gaikwad, and Sashwat Kondepudi. 2014. Probabilistic Inference of Unknown Locations: Exploiting Collective Behavior when Individual Data is Scarce. In *Proceedings of the Fifth ACM Symposium on Computing for Development (ACM DEV-5 '14)*. ACM, New York, NY, USA, 103–112. DOI:https://doi.org/10.1145/2674377.2674387
[11] Joshua E Blumenstock and Nathan Eagle. 2012. Divided We Call: Disparities in Access and Use of Mobile Phones in Rwanda. *Information Technology and International Development* 8, 2 (2012), 1–16.
[12] F. Calabrese, G. Di Lorenzo, Liang Liu, and C. Ratti. 2011. Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing* 10, 4 (2011), 36–44. DOI:https://doi.org/10.1109/MPRV.2011.41
[13] Yves-Alexandre de Montjoye, CÃľsar A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports* 3 (March 2013). DOI:https://doi.org/10.1038/srep01376
[14] Department of Census and Statistics, Ministry of National Policies and Economic Affairs. 2017. *National Accounts Estimates of Sri Lanka for 2016. Department of Census and Statistics market activity in 2013.* Technical Report. http://www.statistics.gov.lk/national_accounts/dcsna_r2/reports/press_note_2016_english.pdf
[15] Dave Donaldson. 2016. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review* (2016). DOI: 10.3386/w16487.
[16] Gilles Duranton and Matthew A. Turner. 2011. The fundamental law of road congestion: Evidence from US cities. *The American Economic Review* (2011), 2616–2652. http://www.jstor.org/stable/23045653
[17] Flavio Finger, Tina Genolet, Lorenzo Mari, Guillaume Constantin de Magny, NoÃńl Magloire Manga, Andrea Rinaldo, and Enrico Bertuzzo. 2016. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proceedings of the National Academy of Sciences* 113, 23 (June 2016), 6421–6426. DOI:https://doi.org/10.1073/pnas.1522305113
[18] Edward Glaeser and Helen Sims. 2015. *Contagion, crime, and congestion: overcoming the downsides of density.* Technical Report.
[19] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (June 2008), 779–782.
[20] John Hine, Maysam Abedin, Tony Airey, Richard Stevens, and Tamala Anderson. 2016. The Poverty Reduction Impact of Rural Roads: A Systematic Review. *EPPI-Centre, Social Science Research Unit Working Paper* (2016).
[21] Hanan G. Jacoby and Bart Minten. 2009. On measuring the benefits of lower transport costs. *Journal of Development Economics* 89, 1 (May 2009), 28–38. DOI:https://doi.org/10.1016/j.jdeveco.2008.06.004
[22] Shahidur R. Khandker, Zaid Bakht, and Gayatri B. Koolwal. 2009. The Poverty Impact of Rural Roads: Evidence from Bangladesh. *Economic Development and Cultural Change* 57, 4 (July 2009), 685–722. DOI:https://doi.org/10.1086/598765
[23] Patrick Kline and Enrico Moretti. 2013. Place Based Policies with Unemployment. *The American Economic Review* 103, 3 (May 2013), 238–243. DOI:https://doi.org/10.1257/aer.103.3.238
[24] Patrick Kline and Enrico Moretti. 2014. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics* 129, 1 (2014), 275–331.
[25] Paul R. Krugman. 1991. *Geography and Trade.* MIT Press. Google-Books-ID: AQDodCHOgJYC.
[26] Rainer Kujala, Talayeh Aledavood, and Jari Saramäki. 2016. Estimation and monitoring of city-to-city travel times using call detail records. *EPJ Data Science* 5, 1 (2016), 6.
[27] A Kumarage. 2012. Sri Lanka Transport Sector Policy Note. *World Bank mimeo* (2012).
[28] Benoit Lefevre, David Leipziger, and Matthew Raifman. 2014. The Trillion Dollar Question. *World Resources Institute* (Jan. 2014). http://www.wri.org/publication/trillion-dollar-question
[29] Maxime Lenormand, Miguel Picornell, Oliva G. CantÃž-Ros, AntÃšnia Tugores, Thomas Louail, Ricardo Herranz, Marc Barthelemy, Enrique FrÃŋas-MartÃŋnez, and JosÃľ J. Ramasco. 2014. Cross-Checking Different Sources of Mobility Information. *PLOS ONE* 9, 8 (Aug. 2014), e105184. DOI:https://doi.org/10.1371/journal.pone.0105184
[30] Xin Lu, Linus Bengtsson, and Petter Holme. 2012. Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences* 109, 29 (July 2012), 11576–11581. DOI:https://doi.org/10.1073/pnas.1203882109
[31] Xin Lu, Erik Wetter, Nita Bharti, Andrew J. Tatem, and Linus Bengtsson. 2013. Approaching the Limit of Predictability in Human Mobility. *Scientific Reports* 3 (Oct. 2013). DOI:https://doi.org/10.1038/srep02923
[32] Xin Lu, David J. Wrathall, PÃěl Roe SundsÃÿy, Md. Nadiruzzaman, Erik Wetter, Asif Iqbal, Taimur Qureshi, Andrew Tatem, Geoffrey Canright, Kenth EngÃÿ-Monsen, and Linus Bengtsson. 2016. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environmental Change* 38 (May 2016), 1–7. DOI:https://doi.org/10.1016/j.gloenvcha.2016.02.002
[33] Carlo Ratti, Dennis Frenchman, Riccardo Maria Pulselli, and Sarah Williams. 2006. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design* 33, 5 (Oct. 2006), 727–748. DOI:https://doi.org/10.1068/b32047
[34] Stephen J. Redding and Matthew A. Turner. 2015. Transportation Costs and the Spatial Organization of Economic Activity. In *Handbook of Regional and Urban Economics*, J. Vernon Henderson and William C. Strange Gilles Duranton (Ed.). Handbook of Regional and Urban Economics, Vol. 5. Elsevier, 1339–1398. http://www.sciencedirect.com/science/article/pii/B978044459531700020X DOI: 10.1016/B978-0-444-59531-7.00020-X.
[35] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabasi. 2010. Limits of Predictability in Human Mobility. *Science* 327, 5968 (Feb. 2010), 1018–1021.
[36] Anthony J. Venables. 2007. Evaluating Urban Transport Improvements: Cost-Benefit Analysis in the Presence of Agglomeration and Income Taxation. *Journal of Transport Economics and Policy* 41, 2 (2007), 173–188. http://www.jstor.org/stable/20054012
[37] Amy Wesolowski, Nathan Eagle, Andrew J. Tatem, David L. Smith, Abdisalan M. Noor, Robert W. Snow, and Caroline O. Buckee. 2012. Quantifying the Impact of Human Mobility on Malaria. *Science* 338, 6104 (Oct. 2012), 267–270. DOI:https://doi.org/10.1126/science.1223467
[38] Amy Wesolowski, Taimur Qureshi, Maciej F. Boni, PÃěl Roe SundsÃÿy, Michael A. Johansson, Syed Basit Rasheed, Kenth EngÃÿ-Monsen, and Caroline O. Buckee. 2015. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences* 112, 38 (Sept. 2015), 11887–11892. DOI:https://doi.org/10.1073/pnas.1504964112
[39] Amy Wesolowski, Gillian Stresman, Nathan Eagle, Jennifer Stevenson, Chrispin Owaga, Elizabeth Marube, Teun Bousema, Christopher Drakeley, Jonathan Cox, and Caroline O. Buckee. 2014. Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. *Scientific Reports* 4 (July 2014), 5678. DOI:https://doi.org/10.1038/srep05678

ACKNOWLEDGMENTS

Kaushalya Madhawa and CD Athuraliya provided excellent research assistance, and Rohan Samarajiva and Saad Gulzar gave valuable feedback on early versions of this work. We gratefully acknowlede support for this research from the International Growth Centre under Grant #1-VRU-VNEC-VXXXX-89117 and from the IZA/DFID Growth and Labour Markets in Low Income Countries Programme under Grant #C2-RA4-205.