Travel time estimation accuracy in developing regions: An empirical case study with Uber data in Delhi-NCR*

Daksh Shah
IIIT-Delhi
daksh17336@iiitd.ac.in

Aravinda Kumaran
IIIT-Delhi
aravinda15133@iiitd.ac.in

Rijurekha Sen
IIT-Delhi
riju@cse.iitd.ac.in

Ponnurangam Kumaraguru†
IIIT-Delhi
pk@iiitd.ac.in

ABSTRACT

Travel time estimates are highly useful in planning urban mobility events. This paper investigates the quality of travel time estimates in the Indian capital city of Delhi and the National Capital Region (NCR). Using Uber mobile and web applications, we collect data about 610 trips from 34 Uber users. We empirically show the unpredictability of travel time estimates for Uber cabs. We also discuss the adverse effects of such unpredictability on passengers waiting for the cabs, leading to a whopping 28.4% of the requested trips being cancelled. Our empirical observations differ significantly from the high accuracies reported in travel time estimation literature. These pessimistic results will hopefully trigger useful investigations in future on why the travel time estimates are mismatching the high accuracy levels reported in literature - (a) is it a lack of training data issue for developing countries or (b) an algorithmic shortcoming that cannot capture the (lack of) historical patterns in developing region travel times or (c) a conscious policy decision by Uber platform or Uber drivers, to mismatch the correctly predicted travel time estimates and increase cab cancellation fees? In the context of smartphone apps extensively generating and utilizing travel time information for urban commute, this paper identifies and discusses the important problem of travel time estimation inaccuracies in developing countries.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

ACM proceedings; LaTeX; text tagging; use semi-colons

ACM Reference Format:

Daksh Shah, Aravinda Kumaran, Rijurekha Sen, and Ponnurangam Kumaraguru. 2019. Travel time estimation accuracy in developing regions: An empirical case study with Uber data in Delhi-NCR. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3308560.3317057

*Produces the permission block, and copyright information

†This work was done in part while Ponnurangam Kumaraguru was on sabbatical at IIIT Hyderabad.

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3317057

1 INTRODUCTION

Travel time estimates are extremely important for urban commuters, for appropriate trip planning. Even in developing countries where transport infrastructure growth is slow, travel time estimates are regularly used for different transport services. In India, for example, public transport fleets are getting fitted with GPS tracking devices and their real time trip data are becoming publicly available [2]. There are also proprietary travel time datasets owned by Google, Uber and similar cab sharing services, crowd-sourced from Google Map users or cab passengers and drivers. All these datasets make travel time estimation possible, based on historical and recent trend analyses using a myriad of prediction algorithms. The estimates are in turn consumed by different transport services: e.g. to estimate the arrival time of buses in bus-stops or cabs at the pick-up locations, or to compute the overall trip times in ongoing cab rides or in Google Maps directions app.

This paper investigates the quality of these travel time estimates in the Indian capital city of Delhi and the National Capital Region (NCR). We use Uber's Estimated Time of Arrival (ETA) of cabs, as a proxy for travel time estimates. In India, the ride-sharing service Uber recorded 1 million trips per day in 2017, including a Delhi commuter taking as many as 5 rides per day [1]. The rate of ridership growth slowed down in 2018, but still there was a significant number of 3.5 million rides per day [7]. Delhi contributes about 10-12% of the Indian trips (about a million trips per week) [6].

Uber's travel time estimation algorithms can therefore use the extensive GPS data that Uber collects from its own cabs. The predictions, in turn, affect the routing and scheduling behaviors of its large and rapidly growing community of passengers and drivers. Thus Uber's travel time estimation numbers form an interesting dataset to empirically examine the travel time prediction quality, as inputs to the estimation algorithms are rich and the outputs are in massive use. Uber acknowledges [9] that its "ETA times are estimates and not guaranteed. A variety of external factors like heavy traffic or road construction can impact travel time." In developing country cities like Delhi-NCR, where traffic congestion and construction are rampant [3–5, 8], this paper explores the quality of Uber's ETA using empirical data from real trips.

State of the art research literature in travel time estimation [13, 14] report around 2 minutes of travel time estimation errors. We, however, find 5-10 minutes of median to 80^{th} percentile errors, with 25 minutes errors in the worst case. These values, empirically computed using 610 trips' data from 34 Uber users in Delhi-NCR, give rise to interesting discussion points in this paper. The errors are not coming from manual reporting errors, as we collect information about users' trips using automated crawling of the Uber smartphone and web apps. Whether the errors come from (i) data or algorithmic shortcomings, or (ii) due to conscious decisions taken by the Uber platform or the Uber drivers, are open questions formulated in this discussion paper, to be explored in future.

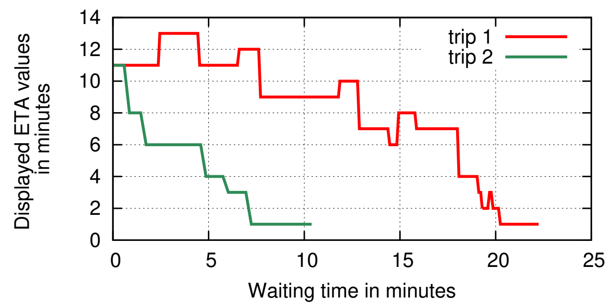


Figure 1: Unpredictable static values and jumps in ETA during waiting, for 11 mins initial ETA

2 UBER'S ARRIVAL TIME ESTIMATES

While assessing the quality of travel time estimates, we take Uber's Estimated Time of Arrival (ETA) for its cabs, as a proxy for travel time estimates. Uber has a web API, which takes as input a GPS coordinate value and returns the Estimated Time of Arrival (ETA) of a cab at that GPS location. This ETA value returned by the API is the same value as shown in the Uber smartphone app (manually verified by us). The Uber user sees this ETA as soon as he opens the app and his location is detected. We refer this ETA as returned by the API and seen at Uber app start time as *t1*.

However, we observe an interesting phenomenon when a booking request is made by the Uber app user. The displayed ETA changes in the Uber app, sometimes showing a significantly different value from *t1*. Our intuitive understanding of this phenomenon is the ETA shown at app start and returned by the API is based on all available cabs in the passenger's neighborhood when the app is started, whereas the value shown after making a booking is based on the actual driver assigned to the request. We denote this ETA shown after a booking request is made as *t2_first*.

t2_first is called so because as the passenger is waiting for the cab, a series of ETA values are displayed in the app, until the cab finally arrives or the trip is cancelled (either by the driver or by the passenger). This range of displayed ETA values as the passenger is waiting is referred to as *t2* and the first ETA as shown in the app is therefore *t2_first*.

Curiously, *t2* remains constant at the same minute over several minutes and also sometimes jumps up instead of monotonically going down, as Uber recalculates the estimations periodically. We refer to time windows during which *t2* remains constant over more than one minute as *t2_stationary* and the instances where *t2* jumps up as *t2_jumps*. As an example, Fig. 1 depicts two different trips with the same *t2_first* along the y-axis as 11 minutes, while the x-axis shows the actual waiting times. Ideally, both curves should monotonically decrease by one minute every minute and reach 0 after 11 minutes. However, while one trip denoted in green as ride 2 has *t2* reaching 0 in about 11 minutes (albeit with few *t2_stationary* instances), the other trip denoted in red as ride 1 has *t2* reaching 0 in almost 21 minutes, with many more *t2_stationary* and *t2_jumps* instances.

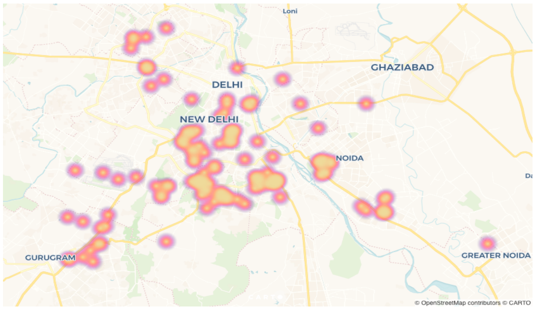
Thus the actual time a passenger has to wait can have low correlation with the estimated travel time displayed in the app. Such unpredictable wait times, where 11 minutes of estimated travel time of the cab to the pickup location can mean exactly 11 minutes or even 21 minutes (91% increase) as seen in Fig. 1, are problematic. It can stress waiting passengers who cannot plan their schedules with confidence, even for important trips. Additionally seeing constant and jumping ETAs displayed in the app while waiting, can increase their frustration. We explore this issue in depth in this paper, by recruiting actual Uber passengers as participants and quantifying the extent to which they experience this unpleasant phenomenon of unpredictable waiting times, due to issues in Uber's travel time estimation.

3 PARTICIPANT RECRUITMENT

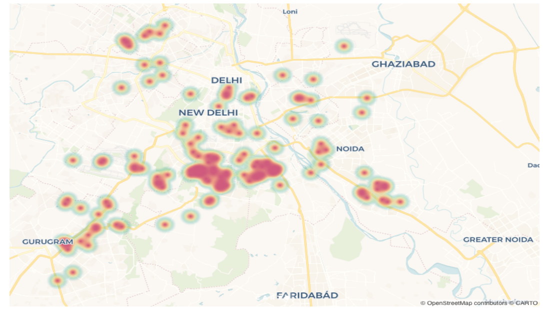
Booking Uber cabs would be unnecessary in this study, if we used the Uber web API to get ETA information. But as mentioned above, we have anecdotal evidence that the ETA shown before and after booking (*t1* and *t2_first* respectively) can vary. We will show empirical evidence of this anecdotal observation in Fig. 3. The web API ETA thus loses meaning as soon as an actual booking is made. Also the web API does not record the actual time of a cab's arrival at the pickup point. Due to these two shortcomings of the web API, we need real bookings to compare Uber's travel time estimations with the actual cab arrival times, and quantify errors if any, for meaningful analyses in this paper.

Booking cabs just for the sake of data analyses would violate Uber's term of services. So we recruit real Uber users in Delhi-NCR and collect ride data using their Uber app authorization tokens. Using personal communication channels like email, social media and word of mouth, we advertised our study and received confirmation from 34 Uber users. They agreed to share their ride data, given that all results that we publish are aggregated statistics without revealing their personal information. These users created their own Uber developer apps using their Uber credentials and shared with us the OAuth (authorization) tokens.

Using these authorization tokens we make Uber API calls to detect if the user is currently requesting a cab, waiting for a cab, taking a ride or has experienced a cab cancellation (done by himself or the driver). We collect this data over two months which gives us 610 unique trips, 437 among them being successful and the remaining 173 cancelled. Fig. 2 gives heatmaps of the pickup and destination locations for the recorded rides, showing their significant geographical coverage, though the participant recruitment through our personal contacts might have had some bias.



(a) **Pickups for Uber trips**



(b) **Destinations for Uber trips**

Figure 2: Heatmaps of source and destination of the trips made by our recruited participants, showing the significant geographical coverage by the participants

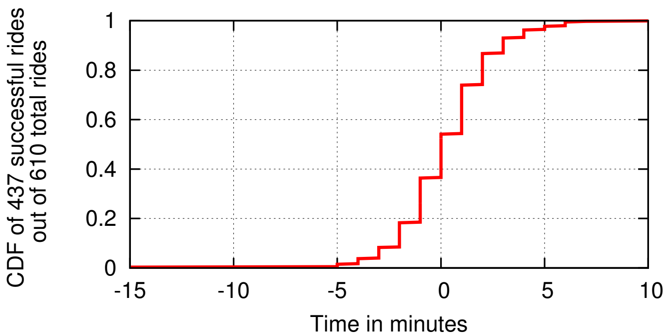


Figure 3: ETA difference before and after booking

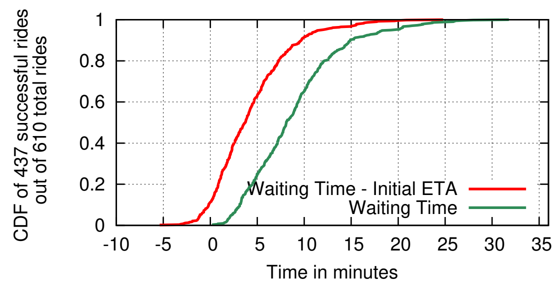


Figure 4: Waiting times and difference between initial ETA displayed after booking and actual waiting times

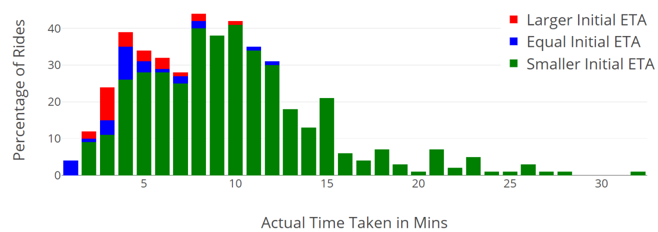


Figure 5: Initial ETA and actual waiting times mismatch

The detailed ETA shown in the app for these 610 trips (*t1*, *t2* with *t2_first* and *t2_last*) and the actual waiting times for the cabs are recorded. The usefulness of participant recruitment is captured in Fig. 3. It plots the CDF of differences between *t1* and *t2_first*, i.e. the ETA displayed before and after booking for the 437 successful trips. This graph computed over the large number of real trips confirm our anecdotal evidence that such difference exist in ETA before and after booking, and therefore motivate our use of real cab booking data to supplement the web API data.

4 PARTICIPANT DATA ANALYSIS

Fig. 4 shows two CDFs, one (in green) depicting the actual waiting time for a cab. It has the top 40^{th} percentile more than 10 minutes for our collected trip dataset. These participants had to wait for more than 10 minutes for their cabs to arrive, which is a significant wait. The median waiting time is 8 minutes.

The second curve in Fig. 4 (in red) shows the CDF of the difference between the ETA shown after booking (*t2_first*) and the actual waiting times before the cab arrives and the trip can start. For less than 20^{th} percentile, the actual waiting times were less than or equal to the ETA, where the curve is to the left of the $x = 0$ line. More common, however, is 5-10 minutes of differences (median to 80^{th} percentile), with the difference going to more than 20-25 minutes in the worst case. Thus the waiting for the cab can be long (more than 10 minutes for top 40^{th} percentile, going upto 35 minutes in the worst case) and there is a gap between the initial ETA shown and the actual waiting time of the user (5-10 minutes differences).

Both these factors of high waiting times and difference between displayed ETA values in the app and the actual waiting times, increase the unreliability of the ride sharing services, as faced by the commuter.

We explore the difference between displayed ETA and actual waiting times in some more detail. Fig. 5 shows the initial ETA after booking along y-xis vs. the actual waiting times along x-axis. It highlights the few instances when the ETA and the waiting times match (in blue), instances where the ETA is an overestimation of the actual waiting times (in red) and instances where the ETA is an underestimation of the actual waiting time (in green). This graph is

Figure 6: Wide range of initial ETA (di erent colors represent di erent initial ETA values) for the same actual waiting time (each bar represents a particular waiting time)

(a) ETA showing constant value for more than a minute

(b) ETA showing upward jumps

Figure 7: ETA remaining constant for more than a minute or jumping to higher values

another depiction of the second CDF in Fig. 4(b). The blue instances with ETA matching the actual waiting times are correct and the red instances overestimating ETA make the users pleasantly surprised when the cab arrives earlier than expected. These are equivalent to the part of the second CDF to the left of the y-axis (x=0) in Fig. 4(b)). The blue and red instances are, however, far less than the green instances of underestimating the actual waiting times, which might make the commuters wary.

Fig. 6 shows a ner granularity of the displayed ETA values along y-axis, with the actual waiting times along x-axis. It further brings out the unreliability of the displayed ETA values. The same actual waiting time can see 9 di erent initial ETA values (each bar showing upto 9 di erent colors, each corresponding to a di erent initial ETA or t2_ rst). The converse way of stating this is the same initial ETA sees a wide range of actual waiting times (each color present in many di erent bars, each bar corresponding to a di erent actual waiting time).

Fig. 1 earlier gave an intuition why the same initial ETA (11 minutes in Fig. 1) led to di erent actual waiting times. During waiting, the ETA remained constant at the same minute value for more than a minute (referred to as t2_stationary above) and also jumped up, instead of monotonically decreasing (referred to as t2_jumps above). We plot these instances of t2_stationary and t2_jumps over our 610 collected trips in Fig. 7. The bars for $x >= 2$ in Fig. 7(a) capture t2_stationary and Fig. 7(b) shows the average and standard deviation of the number of jumps along y-axis vs. actual waiting times on x-axis.

5 BOOKING CANCELLATIONS

173 out of 610 recorded bookings in our study were cancelled (a whopping 28.4%). Fig. 8 shows the last ETA displayed for successful and cancelled trips. While for successful trips the last ETA is predominantly 0 or 1 minute (the cab arrived after that), for cancelled trips this was higher at 7-9 minutes (the trip was cancelled after that). The ne-grained t2_constant and t2_jumps information about the cancelled trips, which cause the ETA to remain high during waiting are also shown in Fig. 7. The frustration of the commuters with Uber's travel time estimates is evident from almost 30% of the bookings being cancelled.

Travel time estimates are meant to pacify commuters, keeping them better informed about what to expect. Uber's erratic travel time estimates and their lack of correspondence with actual cab

Figure 8: Final ETA for successful vs. cancelled trips

arrival times, would however increase commuter stress. Such stress causes the commuters to cancel the cab booking. Cancellations would need them to try and book another cab from Uber, hoping the new assignment will have a short ETA better matching the actual time of cab arrival, or try a di erent ride-sharing company like Ola in Delhi-NCR, or try public transport. Cancellations are also not always free [10 12], so in addition to travel inconvenience, the commuters might also incur nancial loss.

6 DISCUSSION

Based on our empirical analyses, we now formulate some open questions for future work in this discussion paper.

6.1 Gap between literature and practice

The red line in Fig. 4 shows the median 80^{th} percentile di erence between the actual waiting time for cab arrival and Uber's ETA as 5-10 minutes or 300-600 seconds. The worst case di erence is 25 minutes or 1500 seconds. To understand how these numbers compare to the travel time estimation accuracies reported in literature, we refer to two state of the art papers [13, 14].

Travel time estimation algorithms	Porto MAE (sec)	Shanghai MAE (sec)
RTTE [Rahmani et al., 2013]	169.45	214.01
PTTE [Wang et al., 2014]	159.43	168.48
SVR [Asif et al., 2014]	241.41	424.12
SAE [Lv et al., 2015]	222.06	310.47
spd-LSTM [Ma et al., 2015]	217.37	302.45
TEMP[Wang et al., 2016]	193.61	248.70
Deeptravel[Zhang et al., 2018]	113.24	126.59

Table 1: Mean Absolute Error (MAE) of di erent travel time estimation algorithms as given in [14]. We replicate a subset of Table 2. in [14] here.

Using Deep Neural Network based algorithms, the authors in [14] report Mean Absolute Error (MAE) of 113.24 seconds for Porto and 126.59 seconds for Shanghai. Using similar algorithms, the authors in [13] report MAE of 186.93 1.01 seconds for Chengdu and 218.29 1.63 seconds for Beijing. In addition to reporting their own MAE values, the authors also report MAE of other travel time estimation algorithms as baselines.

Table 1 and Table 2 give a subset of these values as given in [13, 14]. As can be seen from the rows in bold in the two tables, the MAE values of state of the art algorithms are far less than 300-600 seconds and worst case error of 1500 seconds, observed over 437 successful Uber trips in Figure 4.

Travel time estimation algorithms	Chengdu MAE (sec)	Beijing MAE (sec)
GBDT	266.15 2.24	393.98 2.99
MlpTTE	265.47 1.53	489.54 1.61
RnnTTE	246.52 1.65	275.07 1.48
DeepTTE	186.93 1.01	218.29 1.63

Table 2: Mean Absolute Error (MAE) of di erent travel time estimation algorithms as given in [13]. We replicate a subset of Table 1. in [13] here.

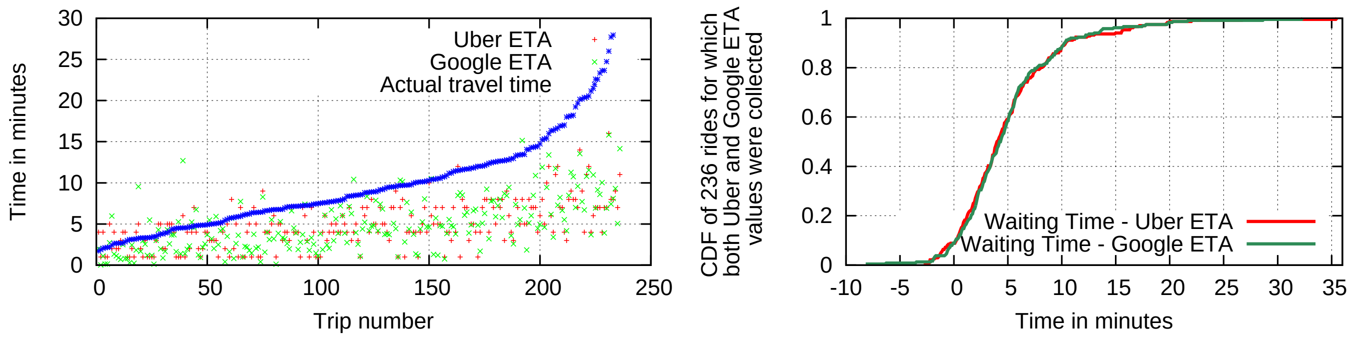
The question is where this gap between Uber's travel time estimates in Delhi-NCR and that in research literature comes from. Is it that Uber does not have enough data to train its algorithms, even though it reports more than a million rides per week in Delhi-NCR [1, 6, 7]? Or is it that the algorithms that Uber uses are unable to capture the (lack of) predictable patterns in Delhi-NCR tra c?

In both Table 1 and Table 2, for the same algorithm in a given row, MAE values di er between two cities of Porto and Shanghai, and that between Chengdu and Beijing. If Delhi MAE were reported in these tables, would the values match our empirical observations of 300-600 seconds MAE? It will be an important direction of future work to examine the limitations of travel time estimation algorithms, in the context of developing countries, to check what MAE the top algorithms give. The challenge will be to get datasets in the scale of Uber or Google, comparable to these urban datasets of Porto, Shanghai, Chengdu and Beijing, to ensure fair comparisons.

6.2 Why is it important to understand the gap

It is important to explore this literature vs. practice gap to check whether the Uber platform intentionally de ates the ETA values or the Uber drivers intentionally mismatch the correctly predicted travel time estimates. These questions arise due to certain characteristic of Uber's cab cancellation policies. In India, if commuters cancel their cab within 5 minutes of booking, they do not incur any cost [10, 11]. However, beyond 5 minutes, the commuters incur a fee. This is the exact opposite cancellation fee policy that Uber has globally [12], where no charges are incurred within 5 minutes of cancellation whereas commuters incur cancellation charges after 5 minutes since booking. This di erence in cancellation policies itself is very interesting. What leads to a di erent policy in India compared to the rest of the world?

In India, therefore, Uber has an incentive to keep passengers waiting for more than 5 minutes, showing them de ated ETA values. The correct high ETA value, if shown to the passenger just after booking, might cause him or her to immediately cancel the booking for free. If Uber intentionally wants to make a pro t out of cancellations, then showing de ated ETA values that remains constant or jumps up for more than 5 minutes, is useful. It is also possible that drivers get a share of the cancellation fees, and it is them who intentionally drive slowly or do not start driving at all, to violate Uber's correct time estimates. It is not the Uber platform, but the drivers who want the commuters to cancel the cabs. We need to verify in future Uber's policies for drivers, whether it is



(a) ETA and actual waiting times, for Uber and Google

(b) Difference between ETA and actual waiting times, for Uber and Google

Figure 9: Google vs. Uber ETA values. Though the absolute values are different in (a), the distribution of the differences with actual waiting times is very similar in (b).

possible for them to monitor such intentional delays on the part of drivers and penalize them when detected.

6.3 Preliminary comparisons with Google data

For the last 236 out of the 437 successful rides, we started collecting the Google travel time estimate from the Uber cab's current location to the cab's pickup location, using the Google Directions API. This gave us the Google ETA, corresponding to the Uber ETA, for these 236 trips. Fig. 9(a) shows the absolute values of the actual waiting times and Uber and Google ETA. Since the Google and Uber ETA values do not completely overlap, this indicates that Uber does not call Google's directions API for ETA. It shows its own computed ETA values.

We compute the differences between both these ETA values and the actual waiting times. Fig. 9(b) shows two CDFs of the differences, which look very similar for Google and Uber. Thus though the two companies compute different ETA values, they both differ similarly with the actual waiting times - 20^{th} percentile overestimation where the actual waiting time is less than or equal to the ETA, median to 80^{th} percentile underestimation of 5-10 minutes and 25 minutes underestimation in the worst case.

Our preliminary investigations indicate that this difference between ETA and actual travel times is probably not intentional on part of the Uber platform. Since the two companies independently give similar ETA difference distributions, then either it is actually hard to estimate the ETA correctly (Google and Uber do not have enough data or the correct algorithms), or they actually are giving a correct estimate which Uber drivers are violating to cause cab cancellations. Unless the limitations of data and algorithms are understood with confidence in this context, it will be hard to do these fine-grained transparency studies, involving multiple stake-holders like commuters, drivers and the ride sharing platform. Any claim against a particular stake-holder will be a serious allegation without proof, which should be avoided through data driven understanding of travel time estimation limits in developing regions.

7 CONCLUSION

This paper identifies an important literature vs. practice gap in travel time estimation accuracy in developing regions using empirical data. In future, we will work on quantifying training data and algorithmic limits for this problem in developing regions, by generating large scale travel time datasets. It is necessary to bring more transparency to complex urban mobility services like Uber, and this discussion paper establishes this necessity.

REFERENCES

[1] Indian Express. 2017. Uber recorded 1 million rides a day in India during 2017, Delhi rider took 5 trips daily. Retrieved Mar 7, 2019 from https://indianexpress.com/article/technology/tech-news-technology/uber-recorded-1-million-rides-a-day-in-india-in-2017-\delhi-rider-took-5-trips-daily-4979437/
[2] IIITD. 2018. Delhi Open Transit Data. Retrieved Mar 7, 2019 from https://otd.delhi.gov.in/
[3] Outlook India. 2018. Delhi Is Planning Congestion Tax On Busy Roads, But Will It Work? Retrieved Mar 7, 2019 from https://www.outlookindia.com/website/story/congestion-tax-dear-mandarins-delhi-isnt-singapore-come-up-with-doable-idea/310363
[4] Times of India. 2018. Centre in poll hurry to kick off 4 infrastructure projects in NCR. Retrieved Mar 7, 2019 from https://timesofindia.indiatimes.com/city/delhi/centre-in-poll-hurry-to-kick-off-4-infra-projects/articleshow/68294617.cms
[5] Time of India. 2018. Traffic congestion costs four major Indian cities Rs 1.5 lakh crore a year. Retrieved Mar 7, 2019 from https://timesofindia.indiatimes.com/india/traffic-congestion-costs-four-major-indian-cities-rs-1-5-lakh-crore-a-year/articleshow/63918040.cms
[6] Economic Times. 2017. Taxi aggregator Uber picks up pace narrowing gap with Ola. Retrieved Mar 7, 2019 from https://economictimes.indiatimes.com/small-biz/startups/taxi-aggregator-uber-picks-up-pace-narrowing-gap-with-ola/articleshow/60049678.cms
[7] Economic Times. 2018. Speed bump: Uber, Ola face sharp slowdown in growth of rides in 2018. Retrieved Mar 7, 2019 from https://economictimes.indiatimes.com/small-biz/startups/newsbuzz/uber-ola-enter-slow-lane-in-2018/articleshow/65927443.cms
[8] Hindustan Times. 2018. Car owners may have to pay to use 21 congested Delhi roads. Retrieved Mar 7, 2019 from https://www.hindustantimes.com/delhi-news/car-owners-may-have-to-pay-to-use-21-congested-delhi-roads/story-J65FdBNPg45FNN61CPgacL.html
[9] Uber. [n. d.]. How do ETAs work? Retrieved Mar 7, 2019 from https://help.uber.com/riders/article/how-do-etas-work?nodeId=6a4dbe13-0a86-4d11-a0ab-cc88f8171fb8
[10] Uber. 2017. Cancellation Fees in Delhi. Retrieved Mar 7, 2019 from https://www.uber.com/en-IN/blog/delhi/cancellation-fee/
[11] Uber. 2017. Cancellation Fees in India. Retrieved Mar 7, 2019 from https://www.uber.com/en-IN/blog/understanding-cancellation-fee/
[12] Uber. 2017. Global Cancellation Fees. Retrieved Mar 7, 2019 from https://help.uber.com/riders/article/am-i-charged-for-cancelling?nodeId=24e75a3b-cf44-44e4-abae-8c2dce3b07a3
[13] Dong Wang, Junbo Zhang, Wei Cao, Jian Li, and Yu Zheng. 2018. When Will You Arrive? Estimating Travel Time Based on Deep Neural Networks. In *AAAI*.
[14] Hanyuan Zhang, Hao Wu, Weiwei Sun, and Baihua Zheng. 2018. DEEPTRAVEL: a Neural Network Based Travel Time Estimation Model with Auxiliary Supervision. In *IJCAI*.