

# Timetable and Headway Adherence Assessment using the Gini coefficient and the Lorenz Curve

NEILA BHOURI<sup>1</sup>, SNEHA LAKHOTIA<sup>2</sup>, MAURICE ARON<sup>1</sup>, GEETAM TIWARI<sup>2</sup>

<sup>1</sup>COSYS-GRETTIA, Univ. Gustave Eiffel-IFSTTAR, 77454 Marne-la-Vallée, France

<sup>2</sup>TRIPP, Indian Institute of Technology Hauz Khas, New-Delhi - 110 016, INDIA

*Abstract:* Adherence to the schedule is of prime importance in public transport. This paper presents a specific application of the Gini coefficient, well known indicator in economics, for the headway adherence assessment. The paper shows that Lorenz curve, which is usually used to define mathematically the Gini coefficient, is a good indicator of the users' waiting time when it is based on the bus schedule. When it is computed on the basis of the ratio of observed headway to the schedule, it is a powerful visual tool that can be used by operators to detect the existence of irregularities on a bus line at a glance. An equation gives, in an idealistic case, the impact of any single traffic disturbance on the Gini coefficient, making this coefficient comprehensive. A detailed analysis is developed, based on the bus proportions according to the headway adherence level. These proportions are obtained from new indices coming from the derivative of the Lorenz curve. The values of these indexes alert the operator of any adherence disturbance. The examination of the Lorenz curve takes more time, but is worthwhile, giving the types of the irregularities. The application of these indicators is carried on real-time data from the New Delhi bus network.

*Key-Words:* -Reliability, Gini Coefficient, Lorenz Curve, Public Transport, Transit, Bus, Regularity, Index

## 1 Introduction

The reliability of bus travel time is an objective of major concern for bus operators and users. For users, bus irregularities are associated with bunching phenomena or large gaps between buses, which results in low attractiveness of the service for transit agencies. However, due to traffic conditions and variability in bus demand, deviations from schedules are unavoidable, leading to an overall decrease in the level of service and capacity. The ability to accurately and effectively analyze various performance measures is fundamental to determining how well the bus service is adhering to its service standards. Understanding and developing methods to assess transit operations performance are not only valuable in identifying potential problem areas along a route; it is also constructive in proposing effective strategies to improve service reliability.

Time reliability measures are related to headway adherence or on-time performance measures. Headway adherence is often measured when the bus service is operating at headway of 10 minutes or less scheduled; while on-time performance is evaluated for all services with a published timetable. In this paper, the reliability of the bus travel time is

studied using the Gini coefficient and the Lorenz curve regardless of the service planning. The Gini coefficient is computed for the ratio of actual (real) headway to schedule allowing the assessment of the respect of the schedule. When the bus service is planned as a timetable, the travel times of buses are converted to headway.

We also show that the Lorenz curve is a powerful graphical tool for the analysis of the bus regularity. When drawn for the scheduled headway only, the Lorenz curve allows the analysis of the regularity of the waiting time of all riders of the bus line and is a good measure from the users' point of view. When it is drawn based on the ratio of the observed to the scheduled headway, it allows analyzing the regularity of the respect of buses of the scheduled headway and is, therefore, a good measure, mainly from the operator's point of view.

The following section of the paper gives a literature review, exposing firstly the research dedicated to the reliability indicators for public transport assessments, and secondly the Gini coefficient use in the transportation field. The third section explains the Gini coefficient and the Lorenz curve. In section 4 we explain that the Lorenz curve based on the scheduled headway is equivalent to the

Lorenz curve based on the riders' waiting time if the headway is established such that to have the same number of riders in each bus; in section 5 we introduce a Gini coefficient based on the {observed/scheduled} headway ratio and we demonstrate that the associated Lorenz curve is a good indicator of the adherence of buses to scheduled headway. Section 6 gives interesting properties of the Lorenz curve and new indexes coming from its derivative. Applications are made on real-time data from 25 bus routes of the New Delhi Indian bus network. Section 7 is dedicated to data processing and headway estimation. In section 8 we discuss different results of the Gini coefficient and Lorenz curve firstly on the route level. Finally, section 9 gives the conclusion and perspectives.

## 2 Literature review

In the literature, reliability measurements are often discussed under three different categories: (1) travel time-based, (2) schedule adherence, and headway based measures:

(1). Concerning travel time-based reliability, firstly [1] defined reliability as "the ability of the service to provide a consistent service over a period of time" he proposed the inverse of the standard deviation of the measure, such as travel times or wait times as indicators.[2] extended this definition by estimating travel time reliability to be measured by the ratio of the mean travel time to the standard deviation. This was to incorporate the differences in the mean values over differing bus routes or sections of a route. [3] measured travel time variation in terms of percentage values of the excess over mean travel time, and [4] measured it as the percentage of times when travel times exceed the mean value by more than 10 percent. [5] defined travel time reliability as the difference between the actual and planned travel time, assuming that the travel times follow a log-normal distribution. [6] defined PT reliability, in terms of travel time variability (TTV) where TTV is a buffer time as presented in [7]. They used the difference between the 90th and 10th percentile of day-to-day travel times. [8] also proposed the use of TTV and lately [9] proposed the use of TTV with the difference between the 85th percentile and 15th percentile in the numerator.

(2). [10] and [11] proposed a schedule to adherence measures to develop a service quality index. [12]

defined reliability in terms of the ratio of the number of on-time arrivals of the buses to the total bus arrivals. Similarly [13] considered the Chicago Transit Authority's definition of 'running time adherence'. It measures the average difference between the actual and scheduled times and expresses it as a percentage value in terms of the scheduled running times. [2] suggested that service reliability should also be defined in terms of the excess of waiting time experienced by the transit users. This measure is also used by Transport for London [14]. The excess wait time is calculated from the difference between the average wait time for the passengers if the buses were running on schedule and the average wait time experienced. Further [15] defined the 'adherence to schedule' measure with the help of two approaches. The first technique measures the 'reliability of runs that come on schedule', which is estimated by the ratio of the number of implemented runs to the total number of runs scheduled for the same time period. The second approach is defined as a measure of 'punctuality which is closer to the definitions used in other research and is calculated by the ratio of the number of on-time runs to the total number of runs. In this study, runs were considered to be punctual up to 1 minute early and 5 minutes late from the specified scheduled time.

(3). [16-17] emphasized the importance of regular headways, particularly for a high frequency of service. They stated that in the case of regular and even headways, the wait times are minimized, leading to an optimal distribution of passengers in the PT. They estimated headways to be irregular as a measure of the average deviations of the actual headways compared to the scheduled headways. Their analysis revealed that there was a higher degree of regularity at the beginning of the line, and this decreased systematically over consecutive stops.

[2](Liu and Sinha, 2007) further stressed that adhering to the scheduled headways minimizes the wait time, and thereby is seen to positively influence patronage. When the bus service is unable to maintain the scheduled headways, bunching of buses occurs, resulting in an uneven and suboptimal distribution of passengers in the buses, and due to overcrowding, having to bypass bus stops. They followed the definition put forth by [18] which measured headway reliability as the ratio of the standard deviation to the mean headway. [19] argued that the regularity of bus service is more important to the users, rather than the schedule adherence, for high-frequency services. In their

proposed measure, they incorporated the concept of coefficient of variation, since it accounted for the differences in the mean values.

Further to these publications, we find many articles dedicated to the applications of the known measurement to public transport networks. [20] gave brief evaluations of 20 indicators sorted by their function (Travel time indicators, schedule adherence indicators, headway regularity indicators and wait time indicators). The paper takes interest in whether an indicator is “traveler-oriented” or not. According to [20] the preferred indicators to use by bus operators are the percentage of buses running on time and excess waiting time. The authors also proposed a new composite indicator named Journey Time Buffer Indicator “JTBI”.

[21] reviewed nine reliability indicators and gave a comparison between them in terms of ease of understanding, accuracy measure, agency comparability, and cost-efficiency, and gave an overall rank for each one of them. [14] compared the state of practice of 12 bus operators who belong to the International Bus Benchmarking Group (IBBG) [22]. They presented four regularity indicators and listed the advantages and disadvantages of each one of them regarding the ease of communication, objectivity, customer representation and the nature of inputs.

[23] reviewed six types of timetable reliability indicators used in railways and compare them in terms of the information provided, the applicability domain (routes, stops, aggregated) and the necessary inputs for each one. A microscopic simulation model is used to study the indicators' robustness. [24] proposed an indicator named « The Reported Waiting Time” which predicts the waiting time perceived by a traveler; this indicator allows bus operators to better understand the concept of waiting time from the customer’s point of view.

[25] (Teng & Lai, 2015) proposed a new formulation of Bus Running Indicator (BRI) based on Bus Planning Travel Time (BPTT) which was also proposed by authors.

Most of the proposed indicators are not expressed on a normalized scale and therefore cannot be used to compare one route with another. Furthermore, in the majority, they are not immediately or intuitively understandable for senior management or non-expert external stakeholders.

#### **The Gini coefficient in the transport field**

The Gini coefficient used in economic studies to measure the inequality of revenues and health among the population is based on the Lorenz curve [26] (OECD, 2010). In the transport sector, we find

a certain number of papers using the Gini coefficient.

[27] adapted the Gini coefficient and Lorenz curve to assess public transport horizontal equity. Horizontal equity is when all the population has equal transit service regardless of the variability of transit needs within population groups. [28] (Ricciardi et al., 2015) compared the public transport vertical equity, using the Gini coefficient (GC), between three vulnerable groups: elderly residents, no-income households, and no-car households. GC has been largely used in the evaluation of public transport equity, in addition to these articles readers are referred to [29] and [30] .

[31] proposed the use of GC as an evaluation of travel time in order to assess its evenness among road users. GC is calculated in a case study of roads in Korea and is compared with standard deviation, speed, buffer time and buffer index to evaluate the significance of this measure; results show that the Buffer index has a higher positive correlation with the GC in this study.

To the best of our knowledge, only [32] proposed a Regularity index based on the Gini coefficient. The authors studied the regularity of scheduled headway. The indicator was described, however, as difficult to understand and to use by the [33] (TCRP, 2013). We hope that, with the new developments made here, -notably helping the comparisons between bus lines -, that this type of criticism will no longer be justified.

We show in this paper that the application of the Gini coefficient to the schedule is of limited interest. The GC is, however, a powerful indicator of regularity when applied to the ratio of observed to scheduled headway. We also show that the Lorenz curve offers good visualization of the bus regularity which could help operators to have a quick synthetic evaluation of the bus route regularity.

### **3 Gini coefficient(GC)**

The Gini coefficient is often used in economic studies [26] (OECD, 2010). It is a relative measure intended to represent the income distribution of a nation's residents. The Gini coefficient of a developing country may increase (due to increasing inequality of income) while the number of people in absolute poverty decreases.

The Gini coefficient is usually defined mathematically based on the Lorenz curve. A Lorenz curve plots the cumulative percentages of total income received against the cumulative

percentage of the population, starting with the poorest individual or household (see Figure 1). The Gini coefficient can be thought of as the ratio of the area that lies between the line of equality (the bisector) and the Lorenz curve (the surface marked A on Figure 1) over the total area under the line of equality (the surface marked B on Figure 1); i.e., the Gini coefficient  $GC = A / (A + B)$ . Thus a Gini coefficient of zero expresses perfect equality, when all values are the same (for example, when everyone has the same income). A Gini coefficient of one (or 100%) expresses maximal inequality among values (for example, when only one person has all the income or consumption, and all others have none).

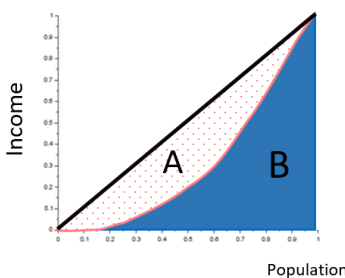


Figure 1: Lorenz Curve Definition

$GC = A / (A + B)$ . Since  $A + B = 0.5$ , the Gini coefficient is  $GC = 2 A$ , or  $GC = 1 - 2 B$ . (1)

If the  $n$  persons ( $i=1..n$ ) of the population are sorted by increasing income  $y_i$ , the total income is  $\sum_{i=1}^n y_i$ . The average income is  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . The coordinates of the points of the Lorenz curve are  $\left\{ \frac{i}{n}; \sum_{j=1}^i \frac{y_j}{n \cdot \bar{y}} \right\}, i = 1, n$ .

The surface over the Lorenz curve (until  $x=1$  and  $y=1$ ) is the sum, from  $i=1$  to  $n$ , of horizontal trapezoids of height  $\frac{y_i}{n \cdot \bar{y}}$  and width  $(i-1/2)/n$ . Then the surface A is equal to that sum, minus 0.5 (the surface of the triangle over the first bisector); then GC is:

$$GC = 2 \cdot A = 2 \left[ \sum_{i=1}^n \frac{y_i}{n \cdot \bar{y}} \left( \frac{i-1/2}{n} \right) - \frac{1}{2} \right] = \frac{2}{n} \sum_{i=1}^n \frac{i \cdot y_i}{n \cdot \bar{y}} - \frac{n+1}{n}$$
 (2)

The averages of the series  $\left\{ \frac{y_i}{\bar{y}} \right\}$  and  $\left\{ \frac{i}{n} \right\}$  being

respectively 1 and  $\frac{n+1}{2 \cdot n}$ , the last term of (Eq.2)

$-\frac{n+1}{n}$  appears as the opposite of the product of these averages (multiplied by 2). Rearranging (Eq.2) gives the GC as a covariance:

$$GC = \frac{2}{n} \cdot \left[ \sum_{i=1}^n \left( \frac{y_i}{\bar{y}} - 1 \right) \left( \frac{i}{n} - \frac{n+1}{2 \cdot n} \right) \right] = 2 \text{cov} \left( \left\{ \frac{y_i}{\bar{y}} \right\}; \left\{ \frac{i}{n} \right\} \right)$$
 (3)

Let us note  $h_i$ , the scheduled headway  $h_i$  of bus  $i$ , and  $h_i + \Delta h_i$  the observed one,  $\Delta h_i$  being the deviation (delay) on the headway.

The sum of headways throughout the day (thus the average headway) does not change in case of disturbances, provided that these disturbances have disappeared at the end of the day(service). This induces an important simplification in the GC calculation and interpretation, and is the rationale for using the Gini coefficient for the ratios

$$y_i = \frac{h_i + \Delta h_i}{h_i}$$
 (see section 5, which is the main contribution of this paper). Indeed, let us take two assumptions: first that this deviation is compensated by the opposite deviation ( $-\Delta h_i$ ) on one of the following bus; second that the scheduled headway of this following bus is also  $h_i$ ; then the deviation on the headway ratio  $\Delta h_i/h_i$  is compensated by the opposite  $-\Delta h_i/h_i$ . So, when a bus is delayed during the day, the sum of the headway ratios does not change (thus their average too, which is normally close to 1 and which is the value  $\bar{y}$  in (Eq.2)). These two assumptions can be criticized as too "idealistic", but they greatly simplify the computation of the impact of a disturbance on the GC, thus they make the GC much more comprehensive:

Without any delay, the Lorenz curve is the first bisector.

In the idealistic case the disturbances address  $k$  ( $k=1..K$ ) bus couples (the leader delayed, one of the followers compensating with a smaller headway ratio). When sorting this  $k$  disturbances by decreasing  $\Delta h_k/h_k$ , the first (the biggest) disturbance pushes the 1<sup>st</sup> couple of buses to occupy the places  $1/n$  and  $n/n$ ... the  $k$ <sup>th</sup> disturbance pushes the  $k$ <sup>th</sup> couple of buses to occupy the places  $k/n$  and  $(n-$

$k+1)/n$ .  
 Whereas the headway ratios  $y_k$  and  $y_{n-k+1}$  of (Eq.2) would be - without the  $k^{\text{th}}$  disturbance,- equal to 1; within it, their values are  $[1-\Delta h_k/h_k]$  and  $[1+\Delta h_k/h_k]$ . The other ratios  $y_j$  are not modified. The impact on GC of the single  $k^{\text{th}}$  disturbance is, following (Eq.2),

$$\begin{aligned} &-\frac{2k \cdot \Delta_k / h_k}{n \cdot \bar{y}} + \frac{2(n-k+1) \cdot \Delta_k / h_k}{n \cdot \bar{y}} \\ &= \frac{2(1-2 \cdot k + 1/n) \cdot \Delta_k / h_k}{n \cdot \bar{y}} \end{aligned} \quad (4)$$

The impact of  $k^{\text{th}}$  disturbance depends on  $n$  (the number of headways considered),  $\bar{y}$  (which is close to 1) and  $\Delta_k/h_k$  (the ratio between the disturbance and the scheduled headway).

#### 4 Scheduled headway regularity coefficient and waiting time

Let us consider the Lorenz curve based on the scheduled headways. The horizontal axis represents the cumulative proportion of buses ordered from buses with the smallest headways to buses having the largest. The vertical axis represents the cumulative headways of the individual buses as they are arrayed on the horizontal axis. The diagonal line is the function that describes perfectly regular service with equal headways for all buses. Substituting the series  $\{y_i\}$  in (Eq.3) by the series of scheduled headways  $\{h_i\}$ , we obtain the Gini coefficient for the scheduled headways:

$$Gini\_S = \frac{2}{n} \cdot \left[ \sum_{i=1}^n \left( \frac{h_i}{H} - 1 \right) \left( \frac{i}{n} - \frac{n+1}{2 \cdot n} \right) \right] \quad (5)$$

$N$  is the number of headway observations,  $H$  is their mean value,  $i$  is the rank of the headway.

##### 4.1 The Gini coefficient for the Scheduled headways

We show that the Lorenz curve based on the scheduled headways is identical to the Lorenz curve plotting the waiting time when the timetable is established such that to get the same number of users in each bus (except for the passengers of the first bus of the day, for which no headway is attached): it is, therefore, a good indicator from the riders' perspective.

Let us assume, for the sake of simplicity, that the scheduled headways between consecutive buses  $i-1$

and  $i$  are independent of the bus stop on the bus route; this means that there is one unique headway  $h_i$  for the bus  $i$  at all the bus stops. Assume that the bus operator builds the scheduled headways such that all buses take the same number of passengers ( $N_p$ ), equal the forecasted traffic demand for the time corresponding to the headway and that passengers go to the stop randomly without expectations of boarding a particular bus. The cumulative percentage of headways or buses (on the X-axis of the Lorenz curve) is thus equal to the cumulative percentage of passengers taking these buses. The mean waiting time of a passenger is equal to half of the headway; the total waiting time of the passengers of each bus  $i$  is equal to the number of passengers by bus multiplied by half of the headway (Eq.6).

$$w_i = N_p \cdot \frac{h_i}{2} \quad (6)$$

Let us re-number the buses, sorting them according to their headway (the bus with the smallest headway becomes the first bus); the waiting time for all passengers of buses 1 to  $i$  are given by (Eq.7):

$$W_{[1,i]} = \sum_{j=1}^i N_p \cdot \frac{h_j}{2} = \frac{N_p}{2} \cdot \sum_{j=1}^i h_j \quad (7)$$

Where  $[1,i]$  is the set of passengers of scheduled headway less or equal to  $h_i$ . The total waiting time for the whole day on the bus line is:

$$W_T = \frac{N_p}{2} \sum_{j=1}^N h_j \quad (8)$$

The cumulative percentages of the waiting time wasted until the  $i^{\text{th}}$  bus, related to the total waiting time, is:

$$W_{[1,i]} / W_T = \frac{N_p}{2} \cdot \sum_{j=1}^i h_j / \frac{N_p}{2} \sum_{j=1}^N h_j = \sum_{j=1}^i h_j / \sum_{j=1}^N h_j \quad (9)$$

The Lorenz curve plotting the cumulative percentages of waiting time against the cumulative percentages of the total population waiting for buses, starting with the smallest waiting time, is thus identical to the Lorenz curve that plots the cumulative percentages of scheduled headways. Thus the Gini\_S coefficient represents exactly the waiting time inequity.

#### 5 Gini coefficient based on the ratio of observed to scheduled headway as an

### indicator of the adherence to the scheduled

Operators already know the irregularity of scheduled headway implied by the timetable they established taking into account many constraints. Operators are more interested in the adherence of the observed headways to the schedule; that adherence can be analyzed thanks to the Lorenz curve based on the ratios {observed/scheduled} headway; the adherence is assessed by the value of the corresponding Gini coefficient, named in this paper Gini<sub>R</sub>. Like other indicators, the Gini<sub>R</sub> must be used carefully by operators -its interest is to prioritize the actions, to highlight the more severe problems. It is well adapted for headway adherence control policies.

The denominators of the ratios used in the Gini<sub>R</sub> are the scheduled headway. When increasing the number of buses, for example, multiplying by two bus frequency (this can be the case when two bus lines have a common trunk and the stops on the common trunk have twice the number of buses than the peripheries), the scheduled headway are divided by two. Assuming the same distribution of the headway deviations the ratios used in the Gini<sub>R</sub> is multiplied by two; this says that the same absolute deviation costs twice as much as relatively to headway. This has a sense when assuming that the doubled frequency is due to twice as many users. However, from the traveler’s point of view, comparing the reliability of two (or more) lines with different frequencies (or average headway, their inverses of frequencies), it is fair to have the same index value, for the same headway deviation Δ*h*<sub>*i*</sub> whatever the average headway; this is obtained by replacing the observed headway *h*<sub>*i*</sub>+Δ*h*<sub>*i*</sub>, (*h*<sub>*i*</sub> being the scheduled headway and Δ*h*<sub>*i*</sub> the observed deviation) in the numerators of the ratios,

by the quantity  $h_i + \frac{\text{Average\_Headway}_1}{\text{Average\_Headway}_0} \Delta h_i$ ;

Average<sub>Headway</sub><sub>1</sub> is the average headway per day on the line is equal to *D*<sub>*ℓ*</sub>/*N*<sub>*ℓ*</sub>, *D*<sub>*ℓ*</sub> being the daily duty and *N*<sub>*ℓ*</sub> the number of headways (i.e. the number of buses minus 1) for line 1 Average<sub>Headway</sub><sub>0</sub> is the average headway of the line with the least frequent time (the highest average headway). It is equal to *D*<sub>0</sub>/*N*<sub>0</sub> lowest number of services is thus *N*<sub>0</sub>+1) among all lines. The ratio

$\frac{\text{Average\_Headway}_1}{\text{Average\_Headway}_0}$  is less or equal to 1. Then,

whatever the line *ℓ*,  $\frac{N_0 \cdot D_\ell}{N_\ell \cdot D_0} \Delta h_i$  is smaller than Δ*h*<sub>*i*</sub>.

So the modified numerators  $h_i + \frac{N_0 \cdot D_\ell}{N_\ell \cdot D_0} \Delta h_i$  remain positive. Then the modified ratios which are used in the Lorenz curves and the Gini coefficients,

instead of being  $\frac{h_i^\ell + \Delta h_i}{h_i^\ell}$ , become:

$$\frac{h_i^\ell + \frac{N_0 \cdot D_\ell}{N_\ell \cdot D_0} \cdot \Delta h_i}{h_i^\ell} = 1 + \frac{N_0 \cdot D_\ell}{N_\ell \cdot D_0} \cdot \frac{\Delta h_i}{h_i^\ell} \quad (10)$$

The multiplication of the ratio  $\frac{\Delta h_i}{h_i^\ell}$  by the

factor  $\frac{N_0 \cdot D_\ell}{N_\ell \cdot D_0} = \frac{\text{Average\_Headway}_\ell}{\text{Average\_Headway}_0}$  has a

standardization effect. This modified ratio is built to ensure that, on average, the same deviation, the same modified ratio whatever the line *ℓ*.

Analyzing and comparing the schedule adherence for different parts of the day (peak /off-peak) during which the average bus frequency varies, also requires similar modifications of the ratios

The passage from the ratios to “modified” ones changes the average ratio  $\bar{R}_\ell$  in  $\bar{R}_\ell^{new}$ , the Gini-R, the Lorenz curve, its derivative. The following equations give the correspondence:

$$\bar{R}_\ell^{modified} = 1 + \frac{N_0 D_\ell}{N_\ell D_0} (\bar{R}_\ell - 1) \quad (11)$$

$$Gini\_R_\ell^{modified} = \frac{Gini\_R_\ell}{1 + \frac{1}{\bar{R}_\ell} \left( \frac{D_0}{D_\ell} \frac{N_\ell}{N_0} - 1 \right)} \quad (12)$$

$$L_\ell^{modified}(x) = \frac{\frac{N_0}{N_\ell} \frac{D_\ell}{D_0} \cdot \bar{R}_\ell \cdot L_\ell(x) + x \left( 1 - \frac{N_0}{N_\ell} \frac{D_\ell}{D_0} \right)}{1 + \frac{N_0}{N_\ell} \frac{D_\ell}{D_0} \cdot (\bar{R}_\ell - 1)} \quad (13)$$

$$\left( L_{\ell}^{modified} \right)'(x) = \frac{\frac{N_0}{N_{\ell}} \frac{D_{\ell}}{D_0} \cdot \bar{R}_{\ell} \cdot L'_{\ell}(x) + 1 - \frac{N_0}{N_{\ell}} \frac{D_{\ell}}{D_0}}{1 + \frac{N_0}{N_{\ell}} \frac{D_{\ell}}{D_0} \cdot (\bar{R}_{\ell} - 1)} \quad (14)$$

$$\left( L_{\ell}^{modified} \right)'(x) \text{ and } \left( L_{\ell}^{modified} \right)'(x) \text{ being the}$$

Lorenz curve based on modified ratios and its derivative for line  $\ell$ .

Historically, the Gini coefficient, the Lorenz curves display the income inequality in a state. Comparing the inequalities between different states by superimposing Lorenz curves is particularly valid when the average income is the same across the different states. Here the comparisons will be particularly valid if the average ratios  $\bar{R}_{\ell}$  are the same or at least similar for all lines  $\ell$ . When different, the  $\bar{R}_{\ell}$  values must be taken into account in the comparisons between lines.

Let us assume that the first and the last bus of the day are on time; this generally happens, even if it is not strictly. Then two cases may happen:

In the case where there is no missing data nor missing bus, the average ratio is generally close to the value 1 for all bus lines. Indeed there is compensation between the observed headway deviation leading to the following conservative

equation: 
$$\sum_{i=1}^N \Delta h_i = 0$$

This does not strictly imply that  $\sum_{i=1}^N \frac{\Delta h_i}{h_i} = 0$ ,

as the denominator  $h_i$  depends on  $i$ . However, the scheduled headway on a line does not change very often along a day, thus this sum is generally close to zero, implying that the average of the ratios "observed/scheduled headway" is close to 1.

All average ratios  $\bar{R}_{\ell}$  are closed to the value 1, thus to each other. Therefore the visual comparison of the different Lorenz curves is right.

- In the case of missing data or missing bus,- for instance,  $k_{\ell}$  missing bus or missing data on the line  $\ell$ , each with a scheduled headway  $h_{\ell}$ , we have:

$$\sum_{i=1}^N \Delta h_i = h_{\ell} \cdot k_{\ell} > 0$$
. The average ratio is perhaps as much as  $1 + k_{\ell}/N_{\ell}$  depending on line  $\ell$ .

## 6 Properties of the Lorenz curve

The Lorenz curve has different properties that make it a powerful graphical indicator [34]. Its different parts may help the operator to identify the headway adherence to the schedule. As we can see in Figure 2, in one glance, the operator identifies:

- At left, the buses with observed headway shorter than scheduled.
- At the center, the buses respecting more or less the scheduled headway.
- At right, the long or very long intervals as compared to the scheduled ones.

1- The Lorenz curve  $L(x)$  is piecewise linear. Its derivative is piecewise constant. Being a cumulative ratio *proportion*, the slope of the straight line between the points is

$$\left\{ \frac{i}{N_{\ell}} ; L \left( \frac{i}{N_{\ell}} \right) \right\} \text{ and } \left\{ \frac{i+1}{N_{\ell}} ; L \left( \frac{i+1}{N_{\ell}} \right) \right\}$$

$\frac{1}{\bar{R}_{\ell}} \cdot \frac{h_{i+1}^{\ell} + \Delta h_{i+1}}{h_{i+1}^{\ell}}$  where  $\bar{R}_{\ell}$  is the average of the headway ratios.

As the ratios  $\frac{h_i^{\ell} + \Delta h_i}{h_i}$  are sorted from the

lowest to the greatest, this slope (i.e. the derivative of the Lorenz curve) is increasing, the Lorenz curve is convex.

Just before a strict increase of the headway ratio, i.e. for buses  $i$  (which corresponds to the bus

proportion  $x_i = i/N_{\ell}$ ) such as  $\frac{h_i^{\ell} + \Delta h_i}{h_i}$  is strictly

lower than  $\frac{h_{i+1}^{\ell} + \Delta h_{i+1}}{h_{i+1}^{\ell}}$ , the derivative of the

Lorenz curve is discontinuous, passing from

$$\frac{1}{\bar{R}_{\ell}} \cdot \frac{h_i^{\ell} + \Delta h_i}{h_i^{\ell}} \text{ at } x_i \text{ to } \frac{1}{\bar{R}_{\ell}} \cdot \frac{h_{i+1}^{\ell} + \Delta h_{i+1}}{h_{i+1}^{\ell}} \text{ at } x_{i+1}$$

Note that  $R_{\max}^{\ell}$  the highest ratio. The highest value of the derivative (slope) reaching the point  $\{1,1\}$  is equal to  $R_{\max}^{\ell} / \bar{R}_{\ell}$



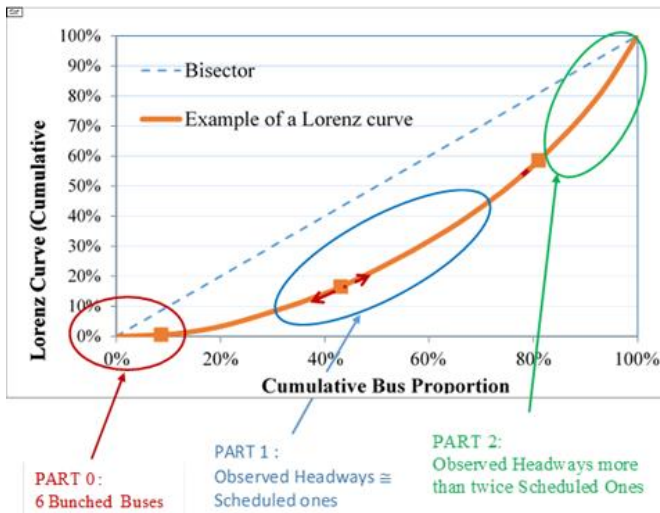


Figure 2: The Lorenz Curve Properties

1- “Part 0”, at the left of the Lorenz curve, gathers the bunched buses (if existing), which have an observed headway near 0, thus a very small headway ratio. Then, the derivative of the Lorenz curve (equal to this very small headway ratio divided by  $\bar{R}_\ell$ ) remains very low..

2- Perfect equity (all ratios equal) lead to a Gini coefficient equal to 0. In this case, Part 1 covers the whole interval [0 1] and the Lorenz curve is confounded with the first bisector. The more problems (whatever their source), the higher the Gini coefficient is and the more the Lorenz curve deviates under the first bisector. “Part 1” is reduced for the benefit of the proportions of low ratios (at the left of Part1) and of high ratios (at the right of Part1).

3- Limits of Part 0, Part 1, Part 2.

Let us define  $x_r$  as the proportion of buses such as the Lorenz curve derivative passes, between  $x_{r-}$  to  $x_{r+}$ , from a value lower than  $r/\bar{R}_\ell$ , to a value equal or greater, i.e. the headway ratio at  $x_{r+1/N_\ell}$  reaches or exceeds the values  $r$ .

$$L'\left(x_{r-}^\ell\right) < \frac{r}{\bar{R}_\ell} \leq L'\left(x_{r+}^\ell\right) \quad (15)$$

The limits of Part 0, Part 1 and Part 2 are obtained by the values of  $x_r$ , for specific values of  $r$ :

Let  $h_\ell$  be the average scheduled headway on line  $\ell$  - for instance  $h_\ell = 14$  minutes. A bus is bunched when its observed headway is only the time of loading/unloading passengers’ -one minute for

instance. Its headway ratio is then below a threshold, obtained by dividing this time by its scheduled headway, *in average*  $h_\ell$  - in the example this threshold is  $r=0.07=1/14$ . Part 0 is then the interval  $[0, x_{0.07}]$ . However it is not excluded that Part 0 includes a few non-bunched buses, when their scheduled headway is much greater errors than the *average* scheduled used for the threshold.

For  $r=1$ ,  $[0, x_1]$  corresponds to buses with an observed headway shorter than scheduled, whereas  $[x_{1+1/N_\ell}, 1]$  corresponds to buses with an observed headway equal or greater than scheduled. Part 1, around  $x_1$ , is defined here as buses whose headway ratio is comprised between  $r=0.75$  and  $r= 1.25$ . This corresponds, for example in the case where the scheduled headway is 14 minutes, to a headway deviation less than 3’30” in absolute value. Thus  $\text{Part1} = [x_{0.75+1/N_\ell}, x_{1.25}]$ .

Part 2” at the right of the Lorenz curve is defined as the proportion of headway ratios greater than 2, which can be considered as irregular.  $\text{Part 2} = [x_{2+1/N_\ell}, 1]$ . The number of such buses is  $N_\ell(1-x_2)-1$ . Their cumulative headway ratio is  $(1-L(x_2)) \cdot \bar{R}_\ell \cdot N_\ell$ . It corresponds to an excess headway ratio  $E_2 = \left[ (1-L(x_2)) \cdot \bar{R}_\ell - (1-x_2) \right] \cdot N_\ell$ .

Headway ratios over 2 occur when:

- a) the previous bus did not pass,
- b) the previous bus passed but the corresponding data was not recorded,
- c) the emergence or aggravation of a traffic jam (or another event) slows down a bus, creating a large headway before the bus,

The number of occurrences of cases a) and b) is equal to the difference between the scheduled number of buses and the observed number of buses in the data file.

Thus  $E_2$ , quantifies the number of missing data, missing bus, cumulative headway deviation (divided by one scheduled headway) for buses whose delay is even more than their scheduled headway. Strictly speaking,  $E_2$  is only a lower bound of this quantity, since a non-recorded bus could sometimes result, behind, in a headway ratio below two: this occurs when the previously recorded bus was delayed.



- Let  $[R_{\max}^{\ell}]$  be the integer part of  $R_{\max}^{\ell}$ . In the case where  $R_{\max}^{\ell}$  is equal or greater than 3, additional  $x_r$  are considered: for all  $r$  integers between 3 and  $[R_{\max}^{\ell}]$ . To make some equations easier, let us complete this series by  $r=[R_{\max}^{\ell}]+1$ , with systematically  $x_{[R_{\max}^{\ell}]+1}^{\ell} = 1$

The interval  $[x_r^{\ell}, x_{r+1}^{\ell}]$  contains  $N_{\ell} \cdot (x_{r+1}^{\ell} - x_r^{\ell})$  buses having an observed headway equal (or greater) than “ $r$ ” times the scheduled one, due either to traffic delays or to “ $r-1$ ” missing data or missing buses, ahead.

The series  $\{(x_{r+1}^{\ell} - x_r^{\ell}), r = 2, [R_{\max}^{\ell}]\}$  constitutes a series of indicators of the worst conditions that we can call “Misery indicators”.

4- When superimposing several lines on the same graphic, the Lorenz curves (and their derivatives) are based on modified headway ratios (the modification depending on the line frequency);

the initial series  $(x_r^{\ell}, r = 0.07 \cdot [R_{\max}^{\ell}])$ , based on non-modified ratios, are identified by the following condition obtained by combining (Eq.14) and (Eq.15):

$$L^{\text{modified}}(x_{r-}^{\ell}) \leq \frac{\frac{N_0 D_{\ell} \bar{R}_{\ell} r}{N_{\ell} D_0} + 1 - \frac{N_0 D_{\ell}}{N_{\ell} D_0}}{1 + N_0 D_{\ell} (\bar{R}_{\ell} - 1) / (N_{\ell} D_0)} = \frac{\frac{N_0 D_{\ell} r}{N_{\ell} D_0} + 1 - \frac{N_0 D_{\ell}}{N_{\ell} D_0}}{\bar{R}_{\ell}^{\text{modified}}} < L^{\text{modified}}(x_{r+}^{\ell}) \quad (16)$$

5 - By analogy with the definition of  $x_r^{\ell}$  in (Eq. 14), let us define  $(x_r^{\ell, \text{modified}}, r = 0.07, [R_{\max}^{\ell}])$  as the proportion of buses such as the derivative of the Lorenz curve for modified ratios  $(L^{\text{modified}})'(x)$  reaches or exceeds  $r / \bar{R}_{\ell}^{\text{modified}}$ . For a given  $r > 1$ ,

this corresponds, for bus number  $N_{\ell} \cdot x_r^{\ell, \text{modified}} + 1$ , to an extra modified headway ratio of  $r-1$ . The extra waiting time for the travelers is, on average only, but whatever  $\ell$ , equal to  $(r-1) \cdot h_0$ ,  $h_0$  being the average headway of line “0”, the line with the lowest frequency. Thus, the proportions  $\{x_r^{\ell, \text{modified}}, \forall \ell\}$  are comparable from the travelers'

point of view and these indicators are powerful. The previous comments related to missing data, missing buses apply on initial  $\{x_{r,\ell}\}$  and not on modified ones.

For  $r=1$ ,  $x_{1+}^{\ell, \text{modified}} = x_{1+}^{\ell}$  since the proportion of buses with  $\Delta h_i \leq 0$  does not change when multiplying

$\Delta h_i$  by  $\frac{N_0 D_1}{N_1 D_0}$  for forming the modified ratio.

## 7 Data

### 7.1 Raw data

Delhi is the capital city of India with a population estimated about 22 million. The mode share of trips by buses is ~20%, with a fleet size of ~10,000 buses plying on ~700 routes [35]. We used the bus data provided by the operator DIMTS for September 2016 for a sample of 25 routes (Figure 1

Figure 1). The AVL data consists of route-geometry specifications, bus stop locations, trip details, and time performance details. These are provided separately for both the directions of each route, classified as “UP” and “DOWN”. Route-geometry data lists the various segments that comprise the route in a particular direction, along with the coordinates of the key locations defining the segment. An example of this file is presented in Table 1. Bus stop location data specifies the sequence of bus stops for the particular direction of the route, and also the coordinates of the stops (see Table 2). The files on trip details and trip performance provide the details of individual bus trips for each day. They also give the information

whether the trip functioned as normal, or was deviated from the route (noted as “route deviated”), or if there was a breakdown and the trip was terminated prematurely (noted as “breakdown”).

Besides, the GPS data from the AVL device on each bus plying on the route was provided. However, these data were not segregated into the two directions of each route. The data contain the details of all the buses running on the specific routes: the bus identification number, the timestamp of when the GPS data was transmitted every 10 seconds, the coordinates of the bus movement for this 10-second frequency, and the instantaneous speed for the corresponding timestamp (Table 3). For our analysis, we considered the route-geometry data, bus stop locations, and GPS data.



Figure 1: Sample of 25 bus routes in Delhi

Table 1: Example of route-geometry data for route 239DOWN

Route name	ID	DO	SL	Geometry (LineString)
239DOWN	0	0	82	LINestring (77.3311810930484 28.6856949227825, 77.331075 28.68609, 77.330955 28.686405)
239DOWN	1	82	35	LINestring (77.330955 28.686405, 77.330834 28.686697)
239DOWN	2	117	119	LINestring (77.330834 28.686697, 77.3304199019109 28.6877081504595)
...	...	...	...	...

**DO: Distance from origin**  
**SL : Segment length**

Table 2: Example of bus stop location data for route 239DOWN

Route name	Stop ID	Stop name	Stop latitude	Stop longitude	Sequence order
239DOWN	U03793	Dilshad Garden Depot (New Seemapuri)	28.6857	77.3312	0
239DOWN	U03789	Shalimar Garden More	28.6877	77.3304	1
239DOWN	U03770	DLF More	28.6929	77.3272	2
239DOWN	DC3794	Tahir Pur	28.6913	77.3249	3
...	...	...	...	...	...

Table 3: Example of bus-based GPS data for route 239

Vehicle Reg No	Date Time Received	Latitude	Longitude	Speed
DL1PC6759	01/09/2016 19:02:51:000	28.6686	77.22654	7
DL1PC6759	01/09/2016 19:03:01:000	28.66853	77.22629	7
DL1PC6759	01/09/2016 19:03:11:000	28.66842	77.2259	14
DL1PC6759	01/09/2016 19:03:21:000	28.66837	77.22572	0
...	...	...	...	...

### 7.2 Travel-time estimation

We compared two different approaches to process the GPS data from the buses to assess the link travel time and speed for the sample of routes. The link is defined as the road section between two consecutive bus stops. In the first approach, we used bus stops as the unit of reference. Buffer areas were created around each bus stop. Speeds were estimated based on the instances of the first detection of the buses in these buffers (approximately as the arrival times of the buses). In the second method, the buses were the unit of reference. Buffer areas were created around the GPS locations of the buses, and travel times (and speeds) were derived from the first instances of when the buses changed the links, i.e. crossed the bus stops (approximately the departure times of the buses).

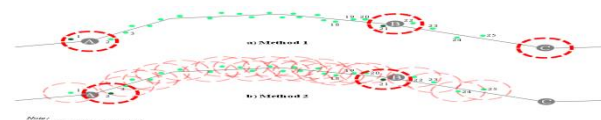


Figure 4: Link Travel-Time and speed estimation

This needed the identification of the bus direction. Note that a U-turn can occur along the route, not necessarily at the terminus. For each GPS coordinate, a link search algorithm was applied in the sequence order of the current route, starting from the first link of the route. This allowed detecting a change in direction when the next link position is found before the previous link position. Once the link position was found, the curvilinear distance from the upstream stop of the link was computed. This distance was also used to anticipate either the detection of a U-turn (when distance decreased) or the arrival at the terminus (when distance equals the link length).

Method 2 differs from method 1 in that method 1 considered arrival times of the buses within the vicinity of the bus stop, and method 2 considered the departure times of the buses from the bus stop. Results showed that differences in the speed estimates by the two methods were statistically significant for only less than ~10% of the links, and there was no significant difference in the headway estimates. Method 2 is however chosen because of the number of missing data, many buses are not detected in the buffer region of the bus stops leading to fewer observations.

**7.2.1 Estimation of link travel times and speeds**

In this method, link travel time was estimated using the following equation:

$$T_i^j = D_e^j - D_s^j \tag{17}$$

Where  $T_i$  is the travel time on the link,  $D$  is the departure time of bus  $j$  from the bus stop,  $e$  is the endpoint (or buffer) of the link, and  $s$  is the start point (or buffer) of the link. It is important to note that the travel time is computed from the difference of the approximate arrival times of the bus at two consecutive stops.

Thus, it is inclusive of the dwell time experienced at the start of the link, i.e. the first stop of the link. Also, speeds greater than 50 km/h were disregarded for the estimates of the mean and SD. This is because link speeds greater than 50 km/h implies even higher instantaneous speeds for buses, which is not feasible. Such kind of estimates is likely results of the errors that arise sometimes in GPS data and are thus discarded.

**7.2.2 Estimation of headways**

The headway  $H_a^{j+1}$  of a bus  $j+1$  at the bus stop  $a$  was estimated using the following equation:

$$H_a^{j+1} = D_a^{j+1} - D_a^j \tag{18}$$

Where  $D_a^j$  is the schedule. As a bus stop may belong to several routes, attention was taken to satisfy this property when computing the headways.

**8- Results**

The Gini coefficient is an inequity measure, however if all buses were late as compared to

their schedule with the same amount of time, the Gini coefficient is equal 0, it is closer to one when there is no similarity at all between the schedule and the observed headways. This helps to identify the most irregular bus routes in a network and the black spots on a bus route.

**8.1- Regularity at bus routes**

*Comparing the respect of the schedule of four different bus routes.*

As shown in (Figure 5) the bus line 165 is the one with the more regular headway, - from the passenger point of view.

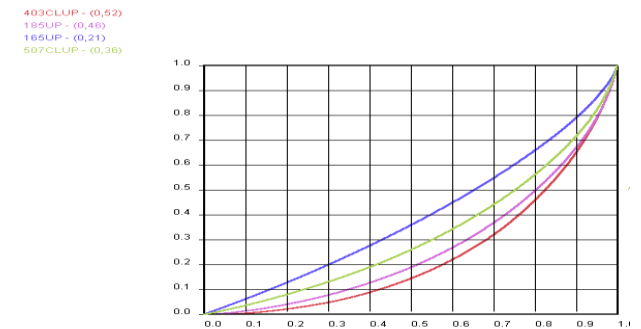


Figure 5a. Direction “UP”

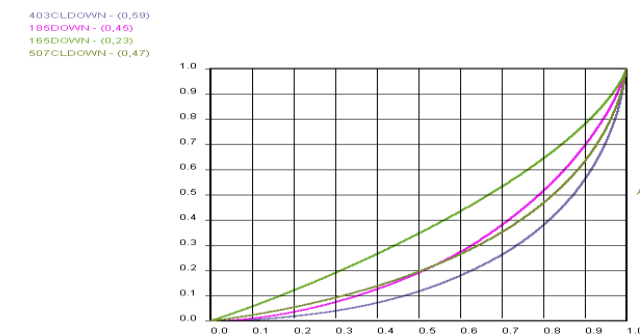


Figure 5b: Direction “DOWN”.

Figure 5: Lorenz curves of four bus routes: lines 165, 185, 403 and 507.

X-axis: cumulated proportion of buses (sorted by increasing observed/ scheduled headway ratio).

Y-Axis: cumulated proportion of observed / scheduled headway ratio.

Considering the number of minutes less or more versus the scheduled headway for the two directions UP and DOWN the modified Gini coefficient is equal respectively to 0.21 and 0.23. The most irregular is the bus route 403 CL with a modified Gini coefficient of 0.52 for the UP direction and a 0.59 for the DOWN. Making a deeper analysis, we notice from the Lorenz curves that bus routes 403CL and 185 have the same amount of long observed headway as compared to the schedule

(common higher part of the Lorenz curve), while the line 403 CL is more bunched than 185 with very short headways, forming what is commonly called a “train of buses”. In the DOWN direction, bus routes 185 23 and 507CL have very comparable regularity (modified Gini coefficients are 0.45, respectively 0.47). The Lorenz curves show occurrences of slightly higher ratios (observed headway longer than scheduled).

**8.2 Regularity at bus stops**

The Gini coefficient is computed at all the bus stops of the different bus routes of our New Delhi sub-network. Figures 6 gives the Gini coefficient at the stops ordered from the origin to the destination of the direction UP (the contrary of Down). It is normal to have differences in the stop names and sometimes on the itineraries of both directions, in that case, we suppressed the corresponding bus stops. We also suppressed bus stops on the two directions when there are missing data in one direction. Results show many missing data at the origin and destination of the bus routes.

For many bus routes, the Gini coefficient tends to increase from the origin to the destination; when superimposing the evolutions of the Gini coefficients of both directions according to the stops, the curve “UP is increasing, whereas the curve “Down” seems to decrease since the order of the stops is reversed (Figure 6 and 7). This finding confirms the observations of [17], who found that the regularity of the line decreases systematically over consecutive stops. However, this is not the case for circular bus routes such as 507CL (Figure 8b)

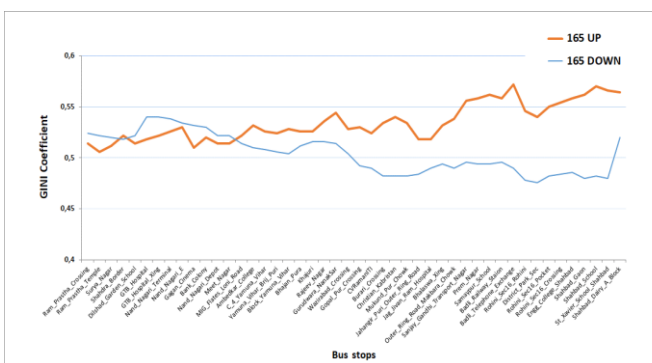


Figure 6: The Gini coefficient value at bus stops in the two directions UP and DOWN for a High Frequency & Long Route – Bus Route 165.

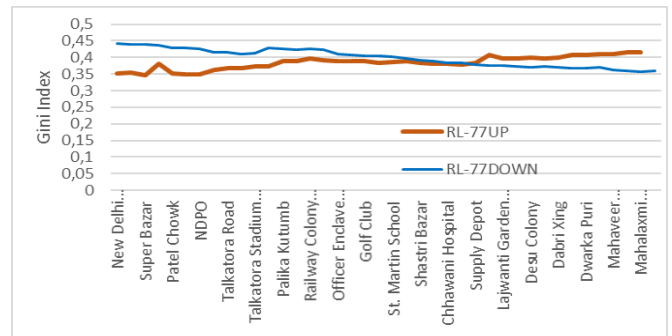


Figure 7: Bus Route RL-77, High Frequency & Short Route



Figure 8.a Map of circular line 507. Both directions

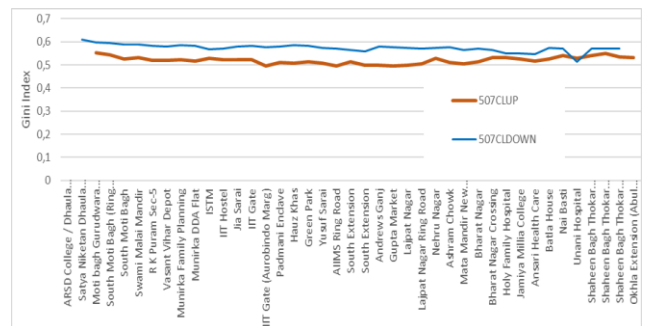


Figure 8b: Circular Line 507CL. Gini by stop

**8.3 The Lorenz curve as a tool of the regularity analysis**

**8.3.1. The example of the Prem-Nagar**

Figure 9 contains the Lorenz curve of the stop Prem Nagar in both directions (UP and DOWN) of line 165 in New Delhi. It is the 41<sup>st</sup> stop of the UP direction and the 12<sup>th</sup> at the DOWN direction.



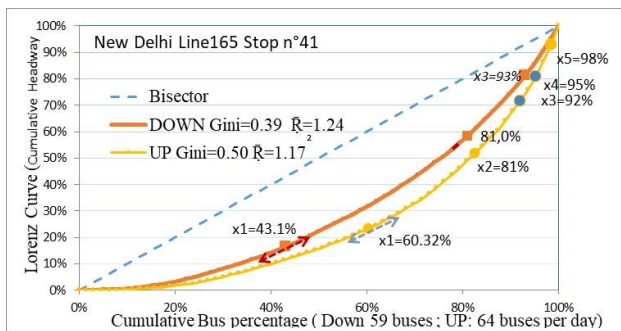


Figure 9: Lorenz curves of {Observed/Schedule} headways at Prem\_Nagar bus stop of bus route 165 for the 2 directions UP and DOWN, the 23<sup>rd</sup> September 2016.

With a higher GINI coefficient in the UP direction, line 165 UP appears less regular than line 165 DOWN at the stop Prem-Nagar, either because congestion or incidents are more frequent in UP than in DOWN, or because the fulfillment of the scheduled headway deteriorates along the line - Prem Nagar is the 41<sup>st</sup> stop in the UP direction, farther from the origin than it is in the DOWN direction.

### 8.3.2 Lorenz Curve Derivative Characteristics at Prem Nagar

The series  $\{x_r\}$  characterizing the Lorenz Curve derivatives for Prem Nagar provides deeper analyses and comparisons between 165 UP and 165 DOWN.

- $[0, x_{0.07}]$  corresponds to Part 0.  $x_{0.07}$  is lower in sense DOWN (3.5%) than in sense UP (10%); referring to the data validates the assumption that Part 0 corresponds to bunched buses -2 bunched
- Between  $x_{0.07}$  and  $x_{0.75}$  are observed headway ratios shorter than scheduled.  $x_{0.75}$  is higher in UP (51%) than in DOWN (32%). Although this is positive for some travelers in UP, we will see below that it is the backlash of difficulties on UP.

- Part 1, from  $x_{0.75}$  to  $x_{1.25}$ , is larger in Down (from 32% to 58%) than in UP (from 51% to 63%), indicating better operations in DOWN

- As  $x_2=81\%$  in both directions, Part 2 =  $]x_2, 1]$  addresses the same proportion of buses (missing bus, missing data, number of buses whose delay has worsened by more than one scheduled headway than the delay of the previous bus.

However the operations are more deteriorated in

the UP direction (incident or congestion), where the higher ratio is more than 5: Indeed the characteristics ( $x_4=95\%$ ) and ( $x_5=98\%$ ) exist only on UP, implying the existence of three additional misery indicators ( $x_4 - x_3$ ), ( $x_5 - x_4$ ), and  $(1 - x_5)$ . This, in turn, implies many observed short or very short headway ratios, for buses passing just after the end of the incidents. That's why the proportion of short headway ratio  $sx_{0.75}$  is higher in UP than in DOWN.

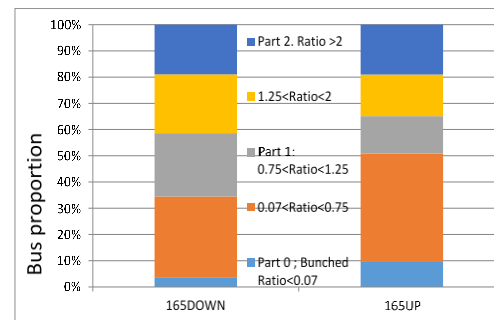


Figure 10: Bus proportions for Part 0, Part 1, Part 2 at Prem Nagar (both directions).

## 9. Conclusion

The Gini coefficient, based on the ratio of observed to scheduled headways is an indicator of the adherence to the schedule; its value is 0 in the case of perfect adherence, thus in case of perfect operations; it is 1 in the most abnormal case of all buses bunched. The operator is alerted by a high Gini coefficient, computed on particular stops or days. The value of the Gini coefficient is comprehensive: an equation gives, in an idealistic case, the impact of any single traffic disturbance on the GC. Comparing different bus lines, or comparing different parts of the day (peak/off-peak.) requires, before computing the Gini coefficient, some homothetic modifications of the ratios in order to accord with the average bus frequencies. We address here the Lorenz curve, which gives the cumulative percentage of ratio spent against the cumulative percentage of buses (ordered by increasing ratios). Then we focus on the derivative of the Lorenz curve, which provides new indicators detailing different proportions of buses with respect to the adherence to scheduled headway: bunched buses, buses with an observed headway close to its scheduled one,

buses with an observed headway twice (or more) than the scheduled one, etc..

When computed on the modified ratios, these new indicators are standardized with regard gain or loss of waiting time. Their values are comparable when applied for different parts of the day or for bus lines with different frequencies.

Once computed, the Lorenz curve helps the operator to identify the regularity of the bus line at a glance. We presented here an application on a part of the New Delhi bus network. We believe that this will complement the knowledge that operators already have on the network they manage.

The perspectives are twofold: (1) addressing the advantages and drawbacks of this coefficient compared with others (standard deviation of headway) – (2) Adding to the observed or scheduled headway, the observed or scheduled travel time from previous to current stop. Then making the ratios and applying the whole process. This would provide a different ranking of buses, suitable for the point of view of the traveler, who is sensitive to his entire journey time.

### References

- [1.] Polus, A. (1978). Modeling and measurements of bus service reliability. *Transportation Research*, 12(4), 253-256.
- [2.] Liu, R., & Sinha, S. (2007). Modelling urban bus service and passenger reliability. *The Third International Symposium on Transportation Network Reliability*, 19-20 July, The Hague, Netherlands.
- [3.] Strathman, J. G., Kimpel, T. J., Dueker, K. J., Gerhart, R. L., & Callas, S. (2002). Evaluation of transit operations: Data applications of Tri-Met's automated bus dispatching system. *Transportation*, 29(3), 321-345.
- [4.] Meyer, M. (2002). Measuring system performance: Key to establishing operations as a core agency mission. *Transportation Research Record. Journal of the Transportation Research Board*, (1817), 155-162.
- [5.] Chang, J. S. (2010). Assessing travel time reliability in transport appraisal. *Journal of Transport Geography*, 18(3), 419-425.
- [6.] Mazloumi, E., Currie, G., & Rose, G. (2010). Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering*, 136(7), 623-631.
- [7.] Bhouri, N., Kauppila, J. (2011) "Managing Highways for Better Reliability – Assessing Reliability Benefits of Ramp Metering, *Transp. Research Record, Journal of the Transportation Research Board of the National Academies*, Washington., D.C., pp. 1-7.
- [8.] Zhi-Sheng Ye & Min Xie (2015), Stochastic modelling and analysis of degradation for highly reliable products. Volume31, Issue1, Special Issue: Actual Impact and Future Perspectives on Stochastic Modelling in Business and Industry. January/February 2015.
- [9.] Glick, T.B., & Figliozzi, M.A. (2017). Measuring the Determinants of Bus Dwell Time: New Insights and Potential Biases. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2647, pp. 109-117
- [10.] Hensher, D. A., & Prioni, P. (2002). A service quality index for area-wide contract performance assessment. *Journal of Transport Economics and Policy (JTEP)*, 36(1), 93-113.
- [11.] Thompson, K., & Schofield, P. (2007). An investigation of the relationship between public transport performance and destination satisfaction. *Journal of transport geography*, 15(2), 136-144.
- [12.] Sheth, C., Triantis, K., & Teodorović, D. (2007). Performance evaluation of bus routes: A provider and passenger perspective. *Transportation Research Part E: Logistics and Transportation Review*, 43(4), 453-478.
- [13.] Lin, J., Wang, P., & Barnum, D. T. (2008). A quality control framework for bus schedule reliability. *Transportation Research Part E: Logistics and Transportation Review*, 44(6), 1086-1098.
- [14.] Trompet, M. & Graham D. (2012) A balanced approach to normalizing bus operational data for performance benchmarking purposes. *Transportation Research Board 91th Annual Meeting*.
- [15.] Eboli, L., Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy*, 18(1), 172-181.
- [16.] Van Oort, N., & van Nes, R. (2007). Improving reliability in urban public transport in strategic and tactical design. In: *Compendium of papers TRB 87th annual meeting*, Washington, DC (pp. 1-17).



- [17.] Van Oort, N., van Nes, R. (2009). Regularity analysis for optimizing urban transit network design. *Public transport*, 1(2), 155-168.
- [18.] Bowman, L. A., Turnquist, M. A. (1981). Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research Part A: General*, 15(6), 465-471.
- [19.] Chen, X., Yu, L., Zhang, Y., & Guo, J. (2009). Analyzing urban bus service reliability at the stop, route, and network levels. *Transportation research part A: policy and practice*, 43(8), 722-734.
- [20.] Gittens, A., Shalaby, A. (2015). Evaluation of bus reliability measures and development of a new composite indicator. *Transportation Research Record: Journal of the Transportation Research Board*, (2533), 91-99.
- [21.] Currie, G., Douglas, N. J., & Kearns, I. (2012, September). An assessment of alternative bus reliability indicators. In Australasian Transport Research Forum
- [22.] IBBG Website (accessed January 2020), <http://busbenchmarking.org/>
- [23.] Jensen, L. W., Landex, A., & Nielsen, O. A. (2014). Evaluation of robustness indicators using railway operation simulation. *Computers in Railways*, 14, 329-339.
- [24.] Fan, Y., Guthrie, A., & Levinson, D. (2016). Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. *Transportation Research Part A: Policy and Practice*, 88, 251-264.
- [25.] Lai, X., Teng, J. (2015.) A Calculation Method for Bus Running Index, *95<sup>th</sup> TRB Annual Meeting*, Washington.
- [26.] OECD, (2010), Income Distribution Database (IDD): Gini, poverty, <sup>2</sup>income, Methods and Concepts. <http://www.oecd.org/els/soc/income-distribution-database.htm>, (accessed January, 2020).
- [27.] Delbosc, A., Currie, G. (2011). Using Lorenz curves to assess public transport equity. *Journal of Transport Geography*, 19(6), 1252-1259.
- [28.] Ricciardi, A.M., Xia, J.C., & Currie, G. (2015). Exploring public transport equity between separate disadvantaged cohorts: A case study in Perth, Australia. *Journal of Transport Geography*, 43, 111-122. <https://doi.org/10.1016/j.jtrangeo.2015.01.011>
- [29.] Jang, S., An, Y., Yi, C., & Lee, S. (2017). Assessing the spatial equity of Seoul's public transportation <sup>2</sup> the Gini coefficient based on its accessibility. *International Journal of Urban Sciences*, 21(1), 91-107.
- [30.] Pavkova, K., Currie, G., Delbosc, A., & Sarvi, M. (2015) A New Approach to Exploring the Operational Performance of Public Transport Links, the case of Melbourne, Australia.
- [31.] Lee, S. B., Lee, S. M., & Lee, K. Y. (2019). A Gini coefficient based evaluation on the reliability of travel time forecasting. *Journal of King Saud University-Engineering Sciences*, 31(4), 314-319.
- [32.] Henderson G., P. Kwong, H. Adkins - 1991 "Regularity indices for evaluating transit performance". *TRR: Journal of the Transportation Research Board*, No 1297, Washington, D.C., pp. 3-9.
- [33.] TCRP, The Transit Cooperative Research Program, (2013) , "Transit Capacity and Quality of Service Manual", *Transportation Research Board of the National Academies, Report 100*, Washington, D.C., 2nd edition.
- [34.] Bhour, N., Aron, M. & Scemama, G. (2016). "Gini Index for Evaluating Bus Reliability Performances for Operators and Riders". *95<sup>th</sup> TRB annual meeting*. p10-14, Washington D.C.
- [35.] Goel, R., & Tiwari, G. (2016). Access-egress and other travel characteristics of metro users in Delhi and its satellite cities. *IATSS Research*, 39(2), 164-172.
- [36.] Lakhota, S., Bhour N., Scemama G., Pahwa A., Totla A., Rao K.R., Lassarre S. & Tiwari G. Comparison of two different methods of estimating headways and link speeds from bus-based GPS data. *JTTE* (accepted)

### Author Contributions:

This work is a collaboration between the four authors.

### Sources of funding

This research is supported by the Indo-French Centre for the Promotion of Advanced Research (IFCPAR/CEFIPRA); as a collaboration project between the French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR), Grettia Laboratory in the Cosys Department at Université Paris-Est France (now included in Université Gustave-Eiffel), and Transportation Research and Injury Prevention Programme (TRIPP), at Indian Institute of Technology Delhi, New Delhi India.