What factors determine the variability of the level of service experienced by users?

Jaime Soza-Parra · Sebastián Raveau · Juan Carlos Muñoz

Abstract This paper explores on new methods for regression models to explain the evolution of headway irregularity among service lines. Coefficient of variation of headways was selected as the independent variable to study because its direct relationship with extra waiting time. In Santiago, Chile, lack of travel time reliability (mostly on waiting time) is one of the main complaints about the public transport system (called Transantiago). To address this issue, several reliability performance indicators were included in the private bus company contracts. Despite these direct incentives, limited noticeable improvements are observed. This work will explore the roots of unreliability for the case of Transantiago, to understand the main causes of its poor performance in this sense. The results should be useful to orient the interventions in the system’s operations, infrastructure and contracts that will improve reliability the most.

Keywords: Headway · Panel model · Public Transport · Reliability · Waiting time
1 Introduction

Understanding how public transport travellers make their decisions (in terms of mode, departure time and route choices) is essential in transport planning. Demand models have been traditionally based on a few significant variables such as fare and travel time. Although usually omitted from planning models, service reliability has been increasingly identified as a key element of travel behaviour (Engelson & Fosgerau, 2016; Fosgerau, 2016). For example, in The Netherlands measures of service reliability have been incorporated in public transport planning models, based on the premise that reliability reveals the difference between travellers’ expectations and experiences (Kouwenhoven et al., 2014). At least three elements are needed to incorporate reliability in a demand model: a monetary value for reliability (VOR), a model that can predict the reliability level of a service based on the context in which it will operate, and a model predicting the marginal impact of reliability indicators on users’ decisions (Kouwenhoven, 2015). For the second element, this is, to predict the reliability level offered by a service based on its context, it is necessary to understand which are the circumstances and variables that affect the level of variability of a public transport service and how they affect it.

In Santiago, Chile, lack of travel time reliability (mostly on waiting time) is one of the main complaints about the public transport system (called Transantiago). To address this issue, several reliability performance indicators were included in the private bus company contracts. Despite these direct incentives, limited noticeable improvements are observed. This work will explore the roots of unreliability for the case of Transantiago, to understand the main causes of its poor performance in this sense. The results should be useful to orient the interventions in the system’s operations, infrastructure and contracts that will improve reliability the most.

In users’ satisfaction, waiting time is known to be weighted more heavily than the time spent inside the vehicle (Raveau et al., 2014). Headways variability affects not only waiting time variability, but also its expected value. Indeed, given a sequence of bus intervals visiting a bus stop, the average waiting time can be expressed as

\[
E(W) = \frac{E(h)}{2} \left(1 + CV(h)^2\right)
\]

where \(E(W)\) is the expected passenger waiting time, \(E(h)\) the mean bus headway, and \(CV(h)\) the coefficient of variation of headways (Osuna & Newell, 1972). If buses visit the stop at regular intervals, this last term would be zero and the waiting time would take its minimum value, i.e. half of the average headway. Thus, the
difference between $E(W)$ and this minimum waiting time is denoted as the excess waiting time and it is due to an unreliable service.

2 Methodology

Headway variability may be affected by many elements of the context in which a bus service operates. Danés (2016) aimed to explain the propagation of this headway variability between consecutive bus stops in Transantiago. The dependent variable selected was the coefficient of variation of headways since it is directly related with excess waiting time. For each bus service operating in Santiago and every stop each of them visits, the CV of the headways over 30-minute periods was computed.

In a dataset with these characteristics there is a strong autocorrelation between independent observations, as the characteristics of a specific bus stop are correlated with the upstream stop’s characteristics. To solve this issue, the headway variability index measured at an upstream stop was also included as an independent variable. This was supposed to allow the model to explain only the headway variability induced within both stops, and not the variability occurring elsewhere upstream in the route. However, when performing a linear regression to obtain the parameters, a unit root was found on the upstream coefficient of variation of headways.

Situations like the one described here occurs frequently in transportation modelling. An example is the work performed by Lin & Bertini (2002), where they aim to predict bus arrival time by formulating a Markov chain model. In this work, however, a different approach is proposed to explain headway variability in terms of a set of attributes of the service.

The dynamic headway variability propagation specification has the following form

$$ CV_{ik} = \alpha \cdot CV_{i(k-1)} + \beta \cdot x_{ik} + \eta_{ik} + \epsilon_{ik} $$

Here, $CV_{ik}$ is the coefficient of variation of headways in bus service $i$ at bus stop $k$, $x_{ik}$ contain all the different explanatory variables, and $\beta$, $\eta_{ik}$ and $\epsilon_{ik}$ are the set of parameters, the set of fixed but unobservable service specific effects, and the error term respectively.

The proposed approach seeks to obtain unbiased parameters is a first-difference linear panel model (Balestra & Varadharajan-Krishnakumar, 1987; Baltagi, 1981).
Panel data econometrics has demonstrated to be a practical tool to solve typical problems associated with data quality and characteristics, such as unobserved heterogeneity by exploiting the multi-dimensionality of the information (Croissant & Millo, 2008).

The independent variables considered in the model are grouped in three categories: street, route and bus characteristics. This information is obtained from Automated Vehicle Location data, Automatic Fare Collection data, and other available sources. Regarding street characteristics along the route, the impact of exclusive lanes and segregated lanes, as well as the number of traffic lights between two consecutive bus stops, were considered. Route characteristics are those related to the service design, considering traveller’s trip length, frequency, distance and number of stops from the head of the service to the stop, bus operator, type of service (express or all-stop), time period, passenger demand, off-board payment stop and route congestion. Bus size was also considered, which determines the number of doors and capacity.

Although the specification has considered CV as the main reliability indicator, it is possible to also calibrate models for other dependent variables, such as the headway variance, standard deviation, and the percentile difference with the mean.

### 3 Preliminary results and ongoing work

The preliminary model was calibrated using the plm R package (Croissant & Millo, 2008). Results show that, as expected, upstream disturbances have a significant effect on the service regularity at downstream bus stops. Besides, the unit root disappeared with this formulation, which may indicate that not considering this effect could lead to erroneously estimate the parameters. Traffic congestion, explained by the mean velocity and its standard deviation respectively, showed to be significant too. Finally, the variables related to special infrastructure, such as segregated corridors and off-board payment zones, exhibited a low significance level due to the presence of endogeneity. By correcting this issue, those variables revealed a significant and positive impact in the performance of the system.

As a further analysis, Arellano-Bond method (Arellano & Bond, 1991) will be carried out to estimate the same parameters of the best model obtained before. Arellano-Bond method is a generalized method of moments which solves endogeneity in the dependent variables without trading off the sample size. This way, it will be interesting to compare both approaches to test the presence of this supposed endogeneity within the data.
To evaluate the forecasting capability of the resulting model, a validation process will be conducted. For this purpose, a subset of the available data will not be used in the calibration process. This validation data will be used to compare the observed and predicted reliability indicators, computing the mean error, mean squared error and the mean absolute error. This process will allow us to assess if this type of models could be used in the evaluation of a public investment project.

The contribution of this work is the calibration of a predictive model, able to estimate changes in reliability of a public transport service, induced by a given project. For example, it will be possible to predict the impact in reliability of segregating a bus lane or equipping a set of bus stops with off-board payment systems. This way, projects with limited impact in average travel time, but which reduce headway variability, may justify their implementation. Currently, there is no methodology available for cost-benefit analysis to evaluate the benefits of a transport project that affects headway variability significantly. Thus, the proposed model will not only be novel, but also useful.

**Acknowledgements:** This research was supported by the Centro de Desarrollo Urbano Sustantable, CEDEUS (Conicyt/Fondap 15110020), the Bus Rapid Transit Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF), and the scholarship funded by CONICYT for Ph.D. studies (CONICYT-PCHA/Doctorado Nacional/2016).

**References**


