Preferences of Public Transit Commuters: Evidence from Smart Card Data in Singapore

Sumit Agarwal[1], Mi Diao[2], Jussi Keppo[3], Tien Foo Sing[4]

3 October 2020

Abstract

This study employs an administrative dataset containing high-frequency transaction records for approximately four million smart transit cards used by Singaporean residents in order to study the travel preferences of public transport commuters. We examine the impact of service attributes, including travel time, reliability and travel cost, on commuters' transportation mode choices and estimate the implied value of travel time (VOT) and value of reliability (VOR) for different types of public transit commuters. The results show significant heterogeneity in the transport preferences of commuters, with adult transit commuters having higher VOT and VOR relative to senior citizens, students and children. Within the adult group, commuters who frequently switch their transportation modes have significantly higher VOT and VOR. These results have important implications for policymakers in formulating strategies to improve the efficiency of the public transport system, and having flexible and customized transport services could increase the utility and satisfaction levels of public transit commuters.

Keywords: *Public transportation, travel preference, value of travel time, value of reliability, big data*

---

1 Department of Finance, Business School, National University of Singapore. Email: ushakri@yahoo.com
2 College of Architecture and Urban Planning, Tongji University, China. Email: diaomi@gmail.com
3 Business School, Institute of Operations Research and Analytics, and Institute of Data Science, National University of Singapore. Email: keppo@nus.edu.sg
4 Department of Real Estate and Institute of Real Estate and Urban Studies (IREUS), National University of Singapore. Email: rststf@nus.edu.sg

We would like to thank the Land Transport Authority (LTA) for sharing the main empirical data and the Singapore-MIT Alliance in Research and Technology (SMART) for additional data support.

1. Introduction

Improving the efficiency of urban transport systems and the well-being of commuters has been a major challenge for transport authorities. Unlike the airline, banking, and telecommunication industries, which provide diversified services to cater to the different needs of consumers, public transport services are undifferentiated in many cities. Users are required to adjust their travel preferences to suit the transport services. The research gap with respect to quality of and the satisfaction level with public transport services has attracted the interest of many researchers.

A series of papers in the US ran congestion pricing experiments using major highways, including State Route 91 (SR91) and Interstate 15 (I-15) in southern California, Interstate 10 (I-10) in Houston, and Interstate 394 (I-394) in Minnesota, to examine the travel preferences of motorists (Lam and Small, 2001; Small and Yan, 2001; Small, Winston, and Yan, 2005; Small, Winston, and Yan, 2006; Liu, Recker, and Chen, 2004; Liu, He, and Recker, 2007). By evaluating motorists' choice with respect to the quality of services between express toll lanes and free, albeit congested, lanes in the experiments, these studies provide valuable insights into the value pricing schemes designed to account for motorists' heterogeneous preferences for travel time and travel time reliability. However, studies on the preferences of public transit commuters, which are useful for assessments of public transport systems, are relatively sparse.

This study evaluates the travel preferences of public transit commuters in Singapore using a large dataset of smart transit card transactions.[1] We examine the mode choices of public transit commuters between buses and rails and estimate the implied value of travel time (VOT) and value of reliability (VOR) by type of traveler. We address key issues regarding urban transit commuters' preferences, which have been neglected in the urban economic literature.

First, based on a high-frequency dataset on the travel activities of approximately four million cardholders over a one-month period, we derive detailed measures of service attributes that influence commuters' travel behavior, which include travel time, travel time reliability, and travel cost. Previous travel demand studies have mainly focused on the trade-off between travel time and cost but neglect the importance of travel time reliability in the mobility choices of commuters (Carrion and Levinson, 2012). Transport models that omit travel time reliability could bias the estimates of the travel time and cost effects, and policies designed without considering service reliability could possibly be misguided (Bhat and Sardesai, 2006). The big data needed to measure service reliability are not easily and readily available for research purposes. Researchers often resort to collecting self-reported responses via surveys and use these responses to infer commuters' preferences for reliability and other trip attributes (Bates, Polak, Jones, and Cook, 2001; Calfee and Windston, 1998; Asensio and Matas 2008). The validity of survey responses has been challenged repeatedly due to discrepancies between stated preferences in hypothetical situations and the actual choices made by the sample households. Other studies use the preferences revealed in motorists' choices of toll lanes from loop detector data (Lam and Small, 2001) or floating car data (Small, Winston, and Yan, 2005; Small, Winston, and Yan, 2006) to derive travel time and

[1] More details on the smart transit card, which is known as an EZ-Link card in Singapore, are available at http://www.ezlink.com.sg

reliability measures. We use detailed trip-level information from the smart transit card data to compute comprehensive measures of travel time, travel reliability, and travel cost for rail transit and buses at different times of the day and for different origin-destination (OD) pairs.

Second, individual heterogeneity plays an important role in transport modeling. Other than the observed socioeconomic characteristics of individuals, we use fixed effects for a large sample of commuters in the model to control for unobservable heterogeneity, such as travel attitudes, income, health, and lifestyle preferences. Third, we address potential endogeneity between individual mode choices and service-level variation by testing commuters' behavioral responses to rainfall shocks via changes in service attributes.

Last, studies examining the revealed mode choice between public transport and private cars do not account for utility differences in transit time for public transit riders when computing VOT. However, given that private cars and public transport are two distinct and segmented markets, the omission of transit time could underestimate public transit riders' preferences in these studies. Due to land scarcity, Singapore's government has adopted a vehicle quota system to regulate the car population. The government curtails private car ownership by imposing high taxes that increase car prices 5 to 6 times above market values. [2] The car ownership rate of 100 cars per 1000 people in Singapore is relatively low compared to other countries with a similar income level; for example, there are 838 cars per 1000 people in the US. However, public transit fares are kept at a low level to encourage more people to substitute cars with public transport. Public transportation constitutes approximately 63% of trips taken by residents of Singapore during the rush hour, based on the Land Transport Authority's (LTA) estimates. Therefore, Singapore's transit-dominated environment provides an ideal setup to investigate the travel preferences of public transit commuters.

Using the smart transit card data, we find significant variations in public transport mode choices across different groups of riders. Increases in travel time and travel cost reduce the probability of choosing an urban transit mode for a journey. Unreliable travel time has a significant negative impact on the transport mode choices of adult riders, who switch frequently between different public transportation modes, but the impact on other passengers is insignificant. After adjusting travel time unreliability for longer trips, the mode choices of all transit commuters except for senior citizens are highly sensitive to travel service reliability. We find that public transport riders are highly sensitive to service reliability for work-related journeys. This implies that the preference for service reliability is dependent on the purpose of the trip.

Small and Yan (2001) demonstrate that accounting for heterogeneity in VOT is important in evaluating policies to improve the efficiency of urban transport systems. We show that the VOT and VOR of adult transit riders are higher than those of other groups of transit commuters, such as senior citizens, students and children. On average, the VOT of adult commuters is estimated at approximately S$2/hour, or the equivalent of approximately 15% to 16% of their hourly wage,

[2] According to Diao (2019), Singapore residents pay approximately S$158k for a car (Toyota Camry 2.0 Auto Petrol) with an open market value of approximately S$25k. In contrast, the transit fare is estimated at approximately S$0.6 to S$0.8/passenger/trip in 2013 [Source: "LTA's Singapore Land Transport Statistics in Brief 2014"].

which falls below the lower end of the 20%-100% range of the gross wage rate estimated for motorists in other industrialized cities (Small, 1992). The VOR of adult commuters is estimated to be S$0.47/hour when measured with the standard deviation of travel time, or the equivalent of 23% of their VOT; when measured with the 50^{th} to 90^{th} percentile travel time range, the VOR is estimated to be S$0.29/hour, or 14% of their VOT.

We find significant variations in VOT and VOR within the adult rider group. Adult commuters who frequently switch their travel modes have relatively higher values of travel time, reliability and the reliability ratio (VOR/VOT) compared to the average adult commuter. The VOT of the frequent mode-switchers ranges from approximately S$3.63/hour to S$4.58/hour, or the equivalent of approximately 24.6% to 34.8% of their median hourly wage; the VOR is approximately 111.9% (with reference to the standard deviation of travel time) and 40.8% (with reference to the travel time in the 50^{th} to 90^{th} percentile range) relative to the VOT. These results imply that the willingness-to-pay for travel time and reliability of public transport riders are significantly lower than those of motorists.

We find significant variation in the demand elasticity across different commuter groups. More specifically, the bus travel time elasticity of the bus share is approximately -1.46 for adult commuters, compared to -0.70 and -0.79 for senior citizens and students/children, respectively. The bus fare (transport cost) elasticity of the bus share is approximately -1.52 for adult commuters, compared to -2.41 for senior citizens. The bus reliability elasticity of the bus share is 0.22 for the frequent mode-switching adult commuters, but the impact is minimal on other commuters. Our results withstood a series of robustness checks. We test for potential endogeneity by using rainfall as an exogenous shock in the mode choice models and find that expected increases in bus travel time reduce bus demand, especially for short distance trips on days of heavy rainfall. These results are consistent with our elasticity estimates.

Our study has significant implications for policymakers in formulating policies that improve efficiency in public transport systems. Using a novel dataset on smart transit card transactions, we are among the first, as far as we are aware, to provide estimates of VOT and VOR for transit commuters, and these estimates are significantly lower than the estimates for motorists reported in the literature. Understanding the VOT and VOR of transit commuters could help transit authorities to adjust transit services to better promote public transport usage. They could use this information to assess how new transit projects could reduce travel time and improve service reliability for commuters. They could design a more cost-effective fare structure by taking into consideration behavioral responses to travel costs among commuters.

The heterogeneity in travel preferences across different commuter groups implies that the "*one size fits all*" style of public transport services does not cater to the varying preferences of transit commuters. Public transport policies targeting travel time and service reliability improvements have a direct impact on adult commuters in their work-related journeys, whereas fare-based transit policies that offer fare discounts will appeal more to cost-sensitive senior citizen commuters. There are benefits to providing more tailored and flexible services, such as on-demand public transit services and surge pricing, to meet the needs of different passenger types, trip origins, trip destinations, etc. For example, one option would be to offer high-price and high-quality transit services to riders with a high willingness to pay for VOT and VOR while keeping affordable regular services for low VOT and VOR riders.

The rest of the paper is organized as follows. Section 2 presents a theoretical model that incorporates service reliability into individual mode choice behavior. Section 3 describes the public transportation system in Singapore and the smart transit card transaction data used in this study. Section 4 presents empirical findings, and Section 5 concludes the paper.

2. Theoretical Model

Let us consider commuter i who maximizes his/her utility from traveling between origin o and destination d at time t. The utility of the commuter depends on the expected travel time, travel time reliability, and fare as follows:

$$U_{o,d}(z(t),t,i) = \alpha_{t,z}+\alpha_{i,z} + \beta_T(t)E\left[T_{o,d}(z(t),t)\right] + \beta_\sigma(t)\sigma\left[T_{o,d}(z(t),t)\right] + \beta_C(t)C_{o,d}(z(t),t) + e_{o,d}(z(t),t), \quad (1)$$

where the binary variable $z(t) \in \{B, R\}$ corresponds to the mode choice, i.e., traveling by either bus (B) or rail (R) at time t; $\alpha_{t,z}$ and $\alpha_{i,z}$ are mode-specific time and traveler fixed effects, respectively; $\sigma[T]$ is the standard deviation of T; $T_{o,d}(z(t),t)$ is the travel time from o to d by either bus or rail with travel starting at time t; $\beta_T(t)$ and $\beta_\sigma(t)$ are presumably negative coefficients on the mean travel time and the standard deviation, respectively; $\beta_C(t)$ is a negative parameter for the cost; and $e_{o,d}(z(t),t)$ is an *i.i.d.* error term with an extreme value distribution (see, e.g., Embrechts, Kluppelberg, and Mikosch, 1997) driven by factors not included in our model. Equation (1) indicates that the commuter prefers a public transport mode that is faster, lower cost, and less uncertain in travel time.

Commuter i maximizes utility by selecting z such that

$$V_{o,d}(t,i) = \max_{z(t)\in\{B,R\}} U_{o,d}(z(t),t,i). \quad (2)$$

Thus, the commuter uses a bus in an OD journey at time t if, and only if, $U_{o,d}(B,t,i) \geq U_{o,d}(R,t,i)$. Since $e_{o,d}(z(t),t)$ has an extreme value distribution, Eq. (2) corresponds to the probability of choosing a bus:

$$P_{o,d}(B,t,i) = \frac{\exp\left(\widehat{U}_{o,d}(B,t,i)\right)}{\exp\left(\widehat{U}_{o,d}(B,t,i)\right)+\exp\left(\widehat{U}_{o,d}(R,t,i)\right)} = \left[1 + \exp\left(\widehat{U}_{o,d}(R,t,i) - \widehat{U}_{o,d}(B,t,i)\right)\right]^{-1}, \quad (3)$$

where $P_{o,d}(z(t),t,i)$ is the probability that commuter i choses $z(t) \in \{B, R\}$ at time t, with an expected utility of $\widehat{U}_{o,d}(z(t),t,i) = \alpha_{t,z}+\alpha_{i,z} + \beta_T(t)E\left[T_{o,d}(z(t),t)\right] + \beta_\sigma(t)\sigma\left[T_{o,d}(z(t),t)\right] + \beta_C(t)C_{o,d}(z(t),t)$.

The logit function corresponding to Eq. (3) is given by

$$\log\left(\frac{P_{o,d}(B,t,i)}{1-P_{o,d}(B,t,i)}\right) = \widehat{U}_{o,d}(B,t,i) - \widehat{U}_{o,d}(R,t,i), \quad (4)$$

where

$$\widehat{U}_{o,d}(B,t,i) - \widehat{U}_{o,d}(R,t,i)$$
$$= \alpha_{t,B} - \alpha_{t,R} + \alpha_{i,B} - \alpha_{i,R} + \beta_T(t)E\left[T_{o,d}(B,t) - T_{o,d}(R,t)\right]$$
$$+ \beta_\sigma(t)\left(\sigma\left[T_{o,d}(B,t)\right] - \sigma\left[T_{o,d}(R,t)\right]\right) + \beta_C(t)\left(C_{o,d}(B,t) - C_{o,d}(R,t)\right).$$

Thus, the difference in the time fixed effect, $\alpha_t = \alpha_{t,B} - \alpha_{t,R}$, the difference in the individual fixed effect, $\alpha_i = \alpha_{i,B} - \alpha_{i,R}$, and the regression slope parameters $\beta_T(t)$, $\beta_\sigma(t)$, and $\beta_C(t)$ are estimated by either a logit regression or, alternatively, a comparable ordinary least squares (OLS) regression.

Our model can also analyze the effect of exogenous shocks such as rainfall on commuters' choices. Heavy rainfall significantly affects bus travel time due to slowdowns in road traffic during rainy days. However, rain does not affect rail travel time. Therefore, by Eq. (3) and Eq. (4), a rainfall shock decreases the probability of taking a bus:

$$\tilde{P}_{o,d}(B,t,i) - P_{o,d}(B,t,i) \approx \left[1 + \exp\left(\widehat{U}_{o,d}(R,t,i) - \widehat{U}_{o,d}(B,t,i)\right)\right]^{-2} \beta_T(t) \triangle T_{o,d}(B,t) = \left(P_{o,d}(B,t)\right)^2 \beta_T(t) \triangle T_{o,d}(B,t), \quad (5)$$

where $\Delta T_{o,d}(B,t)$ is the increase in the bus's expected travel time due to heavy rain and $\tilde{P}_{o,d}(B,t,i)$ is traveler i's probability of choosing a bus at time t for a trip from O to D on a day with heavy rainfall. Since the coefficient on mean travel time $\beta_T(t) < 0$, Eq. (5) implies that the rainfall shock indeed decreases the probability of choosing a bus.

To model commuter behaviors, we define the value of travel time (VOT), the value of travel time reliability (VOR), and the reliability ratio (RR) as follows. The VOT is the marginal rate of substitution between travel time and cost in a traveler's indirect utility function. Thus, it explains the monetary value of a reduction in travel time for a traveler. The VOR measures the willingness to pay for a reduction in travel time variability. RR is simply the ratio of VOR to VOT, which computes VOR as a fraction of VOT. By Eq. (1), VOT, VOR, and RR can be computed as follows:

$$VOT = \frac{\partial U_{o,d}(z(t),t,i)/\partial E\left[T_{o,d}(z(t),t)\right]}{\partial U_{o,d}(z(t),t,i)/\partial C_{o,d}(z(t),t)} = \frac{\beta_T(t)}{\beta_C(t)} \quad (6a)$$

$$VOR = \frac{\partial U_{o,d}(z(t),t,i)/\partial \sigma\left[T_{o,d}(z(t),t)\right]}{\partial U_{o,d}(z(t),t,i)/\partial C_{o,d}(z(t),t)} = \frac{\beta_\sigma(t)}{\beta_C(t)} \quad (6b)$$

$$RR = \frac{\partial U_{o,d}(z(t),t,i)/\partial \sigma\left[T_{o,d}(z(t),t)\right]}{\partial U_{o,d}(z(t),t,i)/\partial E\left[T_{o,d}(z(t),t)\right]} = \frac{\beta_\sigma(t)}{\beta_T(t)} = \frac{VOR}{VOT} \quad (6c)$$

3. Empirical Data

3.1 Public transit system and smart transit cards in Singapore

Singapore is a small island-state in Southeast Asia with a geographical size of 714 square kilometers and a population of 5.5 million. Its government aspires to transform the island into a "*smart nation*". As part of its smart nation initiatives to promote a cashless and a car-lite society, the government plans to develop an efficient public transport network that provides residents with efficient connectivity from homes to workplaces and any other part of the island with the convenience of a cashless payment system.

Singapore's public transport system consists of a road-based bus fleet system and a rail-based rapid transit system (RTS). The RTS includes an interconnected network of mass rapid transit (MRT) and light rail transit (LRT) rail lines and stations distributed across the island. There are 142 rapid transit stations serving five MRT lines and three LRT lines, supplemented by more than 4,000 bus stops to provide seamless connectivity in the system. The total length of the rail network is 178 km. In 2012, the daily trips by MRT, LRT, and bus reached approximately 5.5 million trips, accounting for approximately 44% of total daily trips (including trips by goods vehicles, public transports, taxis and personal cars) made by Singaporean residents.[3]

In recent years, Singapore's public transport system has become increasingly crowded due to increased ridership, particularly during rush hours. Increased travel demand by residents and foreign tourists and visitors and demand spillovers induced by the government's efforts to reduce private car dependency on the island are potential factors that cause the demand crunch in the public transport system. To alleviate the demand crunch in the RTS, Singapore's government, via the Land Transport Authority (LTA), plans to double the rail transit track length to 360 km by the year 2030. When the RTS network is fully developed, 80% of households will live within a 10-minute walk to the nearest train station.[4] The expanded rail transit network has greatly reshaped the travel behavior of Singaporean residents, which has also caused significant housing price appreciation along rail transit corridors (Song et al., 2020, Zhu and Diao, 2016; Diao et al., 2017). In addition to rail transit investments, LTA plans to improve bus service reliability by expanding the bus fleet and adding more designated bus lanes during peak hours to reduce delays caused by traffic congestion.

In April 2002, a fully automated fare collection system was implemented for public transport, including via MRT, LRT, and bus. Cashless fare payments are facilitated using stored-value contactless (smart) cards known as EZ-link cards. EZ-link cards are accepted for payment both at RTS stations and on public buses. Commuters tap their EZ-link cards on the designated readers when they enter and leave MRT/LRT stations and when they board and alight buses. Commuters who choose to make cash payments pay a higher rate compared to the fare paid via EZ-Link cards for the same journey. EZ-link cards cover nearly 96% of the public transportation trips in Singapore (Prakasam, 2009). There are three types of cards issued: adult, senior citizen, and student/child. Senior citizens and students/children pay concessionary fares for their trips.

The history of smart transit cards can be traced back to 1968 when two German inventors, Dethloff and Grotrupp, developed a plastic card containing a microchip (Shelfer and Procaccino, 2002). Since 1990, the use of smart transit cards has increased dramatically across the world (Blythe, 2004). Data collected from smart transit card transactions, which include detailed records

3 Taxis run by various private operators provide an alternative transport choice for commuters; and the supply of taxis are inelastic in Singapore (Agarwal, et al., 2015; 2019).

4 Source: Land Transport Authority (LTA) of Singapore

on the spatial and temporal information of transit riders, have been increasingly used for transport research. Transport researchers applied smart transit card data to study a myriad of research topics, ranging from transit loyalty (Bagchi and White, 2004, 2005) and individual travel patterns (Kitamura et al., 2006, Morency et al., 2007; Lee and Hickman 2011), to job-housing balances (Huang, et al., 2018). However, modeling individual travel behavior using smart transit card data has yet to be fully explored (Pelletier et al., 2011).

We use an administrative dataset of EZ-link transaction records for all card holders in Singapore for the month of August 2013 to study the travel preferences of public transit riders. The data contain 175,154,156 trips from 3,902,005 cards during the month. The data contain the trip date, card ID, card type, tap-in and tap-out times, boarding stop/station, alighting stop/station, transit fare, bus service number, and vehicle ID for the bus trip. The study period coincided with the normal school terms of public schools in Singapore.[5] All trips in a journey are identified by a unique journey ID. The journeys are divided into single-trip journeys, which are defined as a one-way trip from one location to another, and multiple-trip journeys, which include consecutive trips connected by short transfers between the trips. We estimate the mode choice model (rail versus bus) for public transit riders at the journey level.

In the sampling process, we filter and retain only bus stops that are located within 215 meters (m) of a rail station.[6] This process reduces non-randomness in the distributions of bus stops and rail stations, which may otherwise bias commuters' choice of transport mode for their journeys. We use the 215 m buffer zone from a rail station at the origin (or destination) to select only journeys starting (or ending) in these zones. We also use other cutoff distances—150 m, 180 m and 200 m—to define the buffer zones around rail stations in the robustness checks. Journeys with a travel time longer than 150 minutes and intra-buffer zone journeys are excluded from the sample. After filtering, we retain a total sample of approximately 42 million journeys. Due to the large sample size, we use a 10% random sample from the filtered journeys to calibrate our models.[7] We identify journeys made by frequent transit commuters who made at least 30 journeys in a month, as these frequent commuters are less likely to have other outside options (such as taxis and private cars). We perform a robustness check using infrequently transit commuters. We drop journeys that involve a combination of bus and rail, which account for only a small portion of the total journeys in the sample. After the filtering process, we retain a total sample of 175,503 journeys for our analyses.

Based on the filtered sample journeys, we study commuters' public transport choice behavior and estimate the implied VOT and VOR. We examine heterogeneity in preferences for public transport services by different commuter types as identified by the EZ-link card types

[5] The vacation periods for public schools in Singapore ended on June 30, 2013 (source: http://time.sg/school-holidays-2013).

[6] Based on the median distance to the nearest bus stops of approximately 215 m, we select sample bus stops that located within the buffer zone at each pair of origin and destination (od pair), and the filtering out of bus stops located outside the buffer zones could remove potential selection bias at the start of the journey.

[7] The randomly selected 10% samples have characteristics similar to those of the full sample. Given the large number of observations in the subsamples, the estimated results could only be biased downward, if at all.

(child/student, adult, and senior citizen). The empirical results are discussed in the subsequent sections of the paper.

3.2 Travel time and cost

Commuters face a mode choice problem at the start of a journey. Conditional on choosing either a bus or a rail, we estimate the expected travel time, travel cost, and travel time unreliability for each transport mode chosen. Based on their previous travel experiences, we assume that commuters are informed of the travel cost and the first two moments of travel time for a public transit mode across different time segments during weekdays and weekends and for all the origin-destination (OD) pairs in their journeys. We use all the journeys of transit commuters in August 2013 to compute the means and the standard deviations of travel times in a 30-minute time segment during the weekdays and on weekends.

The tap-in and tap-out records of EZ-Link cards capture the overall travel time for MRT rail journeys. For bus journeys, the data capture only the in-vehicle travel time and omit the waiting time for bus arrivals. The waiting time for buses is an important factor determining service quality (Amin-Naseri and Baradaran, 2014; Keyhani, Schnee, and Weihe, 2017) and, if neglected, could bias the commuters' choice of the bus mode for a journey upward. As the bus arrival and the waiting time at bus stops are not directly observable, we estimate the waiting time based on the assumption that passengers arrive randomly at a bus stop and take the first bus that arrives. More specifically, we assume that there is always room in a bus for commuters and that the arrival time of commuters at a bus stop is uniformly distributed. These assumptions are reasonable in Singapore's context. Buses are an important part of Singapore's urban transport system, with a daily mode share of approximately 25.6% in 2012. Bus services are frequent, and bus operators provide only information on the frequency of bus arrivals and not the detailed bus schedule to the public. Real-time bus arrival information collected via the internet and mobile phones was not available in 2013 when our data were collected. Therefore, bus commuters do not have access to real-time bus arrival information during the study period.

Let us denote the time interval between two consecutive bus service arrivals by T. We estimate T for each bus service in a 30-minute time segment separately in our data set during weekdays and on weekends. The average waiting time for a bus service at a selected bus stop is given by:

$$t_{bus,\ waiting} = 0.5 \cdot T.$$

By the uniform distribution of the arrival time of commuters, the standard deviation of the bus waiting time is given by:

$$\sigma_{bus,waiting} = \sqrt{\frac{1}{12}T^2} .$$

Then, the expected total bus travel time and the standard deviation are given by:

$$t_{bus} = t_{bus,waiting} + t_{bus,in-vehicle} \text{ and } \sigma_{bus} = \sqrt{\sigma^2_{bus,waiting} + \sigma^2_{bus,in-vehicle}} .$$

3.3 Summary statistics

Figure 1 plots the kernel densities for the key variables in deciding between a bus and a rail. Table 1 presents the summary statistics for the variables in our model. Overall, the bus mode constitutes 12.4% of the market share, which is significantly lower than the rail's market share. The mode share differs substantially across the three commuter types. Only 11.0% of adult journeys are by bus, compared to 29.1% and 15.1% of the journeys by senior citizens and students/children, respectively. The bus mode has a longer travel time and lower reliability than the rail mode. On average, the bus travel time is 9.258 minutes longer than the rail travel time for the same journey, and the standard deviation of the bus travel time is 4.095 minutes longer than that of the rail travel time. The difference in the standard deviation of travel time between the two modes is approximately 44.2% of the average travel time difference. The difference between the travel costs is small. On average, the bus fare is S$0.017 lower than the rail fare. Adult commuters make more frequent trips to the central business district (CBD) than other commuters. We use rainfall on rainy days (Table 1) as an exogenous shock to test shifts in the bus mode choice in one of the model specifications.

[Insert Table 1 and Figure 1 here]

4. Empirical Results

Using Equation (4), we calibrate the model of commuters' mode choice behavior as follows:

$$Y = \alpha + \beta_1 \Delta(\text{Travel time}) + \beta_2 \Delta(\text{Travel time unreliability}) + \beta_3 \Delta(\text{Travel cost}) + X' \gamma + \lambda + \tau + \epsilon , \tag{7}$$

where the dependent variable Y is a binary variable that equals 1 if an individual takes a bus for a journey and 0 if he/she chooses a rail. The main explanatory variables are the differences (bus minus rail) in the expected travel times, the travel time uncertainties (as measured by the standard deviations of the travel times) and the travel costs for the journey. γ is a vector of coefficients on the control variables that influence the individuals' mode choices. λ represents the card (commuter) fixed effects, τ represents the day fixed effects, and ε is an *i.i.d.* error term. The card and day fixed effects control for unobserved factors such as commuters' travel attitudes, income, health, and lifestyle preferences, as well as differences between normal weekdays, weekends, and holidays. We estimate Model (7) as an ordinary least squares (OLS) regression, which is more flexible when using many fixed effects (card and day fixed effects). Logit regressions with many fixed effects are prone to incidental parameter problems (see, e.g., Lancaster, 2000). We run the logit regressions as a robustness check, and the results are not significantly different from the results of the OLS models.

4.1 Baseline model

In the baseline model, we include only differences in travel costs, travel times, and travel time unreliability between the bus mode and the rail mode, together with the card (individual commuter) and day fixed effects. We estimate Model (7) for all passengers and separately for adults, senior citizens, and students/children to test for heterogeneity in preferences over service attributes across different groups of commuters as identified by their card type.

To test for heterogeneity within the adult group, we further classified the adult passengers into three subgroups according to their mode-switching strategies. Group 1 uses multiple modes for at least one OD pair, Group 2 uses multiple modes but always stays with one single mode for each OD pair, and Group 3 uses one single mode for all the ODs. We expect Group 1 commuters to be the most sensitive to travel time and travel time unreliability among the adult commuters because they actively switch their mode choices to achieve an optimal outcome. To avoid endogeneity, we use the mode-switching behavior of the adult passengers in the first three weeks of the month to assign the adult passengers into one of the three subgroups and use observations from the rest of the month to calibrate the mode choice models. We estimate the same model specification for adult commuters in Group 1 and Group 2. Group 3 commuters with no mode choice variations are less likely to respond to policy interventions. The descriptive statistics for the three groups of adult commuters are reported in Table 1.

The results of the baseline model are presented in Table 2. The estimated coefficients generally fit the predictions of the theoretical model. The coefficients on travel time and transit cost differences are all negative and mostly significant in all the models, suggesting that longer travel times and higher travel costs decrease the chance of commuters choosing a public transit mode for the same journeys. The travel cost coefficient is insignificant for both Group 1 adult commuters and student/child commuters, although the reason for this result could be different for each of the two groups of commuters. As the travel costs of student commuters are normally covered by their parents, they could be less sensitive to transit fare. Group 1 adult commuters are more sensitive to travel time and less sensitive to travel cost, as demonstrated in subsequent analyses.

The coefficients on the difference in the standard deviation of travel time are negative but statistically insignificant in the adult and the student models. Among the adult commuters, travel time unreliability adversely impacts Group 1 adult commuters' mode choice decision, and this effect is significant at the 0.01 level; however, the impact on Group 2 adult commuters is insignificant. One possible explanation is that adults face a huge "penalty" ("opportunity cost") for not arriving to work and related destinations on time (Small, 1982; Mannering and Hamed, 1990; Noland and Small, 1995; Noland, Small, Koskenoja, and Chu, 1998). Group 1 adult commuters are more sensitive to travel time unreliability than other adult commuters, as we expect because they actively adjust their mode-switching strategies to avoid these potential penalties. We further analyze the job-trip related journeys to the CBD in Section 4.6.

[Insert Table 2 here]

4.2 Robustness and heterogeneity tests

In this section, we perform a set of tests to check the robustness of our findings and to test for heterogeneous effects, including models using alternative unreliability measures, models using data based on alternative cutoff distances from rail stations, the mode choice behavior of commuters without outside options, the mode choices of infrequent transit riders, and the mode choice behavior of commuters in different submarkets. We also calibrate the logit specifications of the baseline models.

4.2.1 Alternative unreliability measures

We use the standard deviation of travel time as the measure of unreliability in the baseline models. In their widely cited paper on VOT and VOR for motorists, Brownstone and Small (2005) use the upper tail of the travel time distributions (e.g., the difference between the 50th and 90th percentiles of travel time) to estimate unreliability using logit regressions, and they estimate the VOR based on the idea that commuters are most averse to being substantially later than expected. To test for the robustness of our findings to the use of alternative unreliability measures, we rerun the baseline models using the difference between the 50th and 90th percentiles of travel time in place of the standard deviation of travel time as the independent variable. The results are reported in Table 3.

The models estimated using these alternative unreliability measures produce similar results to those of the baseline model using the standard deviation of travel time. The coefficients on the travel time difference and the travel cost difference show patterns that are identical to those reported in the baseline model, with only small changes in the magnitude of the estimated coefficients. Travel time unreliability, as measured by the difference between the 50th and 90th percentiles of travel time, is only significant for Group 1 adult commuters who frequently switch their modes of travel, which is the same as the result reported in the baseline model. However, the new unreliability measure has a smaller magnitude for the Group 1 adult passenger model than that of the standard deviation of travel time in the baseline model. A 10-minute increase in the difference between the 50th and 90th percentiles of travel time for a given mode reduces the probability of choosing that mode by 2.6 percent , compared to the 7.2 percent decrease in the probability reported in the baseline model. Cultural differences between Singaporean commuters and US commuters could explain the differences in commuters' aversion to extreme delays (Brownstone and Small, 2005). To make our results comparable to those of Brownstone and Small (2005), we calibrate logit regressions using the alternative measure of unreliability and calculate VOT and VOR. The results are reported below in Table 4.

[Insert Table 3 here]

4.2.2 Different cutoff distances from rail stations

In the baseline models, we use 215 m as the cutoff distance to demarcate areas where commuters have easy accessibility to rail stations and bus stops within walking distance. To check the robustness of the cutoff distance assumption, we use different cutoff distances of 150 m, 180 m and 200 m to rail stations to filter the data and then rerun the models. The results are reported in Table A-1 of the online Appendix.

The main findings still hold in the models using the alternative cutoff distances. However, we observe notable differences in the results of the alternative models and the baseline models. In the baseline model, travel time unreliability is only significant in the Group 1 adult commuter model. As we reduce the buffer size from the rail stations, commuters in different subgroups become more sensitive to the unreliability measure. In the models with the smallest buffer size of 150 m, the adult group shows significant aversion to travel time unreliability at the 0.05 level. The child/student group has a negative coefficient that is marginally significant at the 0.1 level. For Group 1 adult commuters, the negative effects of estimated travel time unreliability on mode choice probability generally increase with the cutoff distance to rail stations. A 10-minute increase in the standard deviation of travel time for a given mode reduces the probability of choosing that mode by 5.5, 5.3, 6.2 and 7.2 percent when the cutoff distance is 150 m, 180 m, 200 m, and 215 m, respectively.

4.2.3 Mode choice of infrequent transit commuters

Our baseline models are estimated based on a sample of frequent transit commuters who have made at least 30 transit journeys during the month. These commuters are less likely to have outside options (such as private cars, taxis, etc.) compared to infrequent transit commuters. To compare differences in preferences over service attributes among commuters having different transit trip frequencies, we run the baseline models using the sample of infrequent transit commuters and report the results in Table A-2 of the online Appendix.

We find that the mode choice behaviors of infrequent transit commuters are not significantly different from those of frequent commuters. The coefficients on the travel time variable are significant and negative in all the subgroups, implying that travel time generates disutility among all infrequent transit commuters. Travel costs reduce the probability of choosing a given transit mode among all the infrequent transit commuters except for children and students. The impact of travel time unreliability is negative but insignificant for all the subgroups of infrequent transit commuters.

4.2.4 Mode choice of transit commuters without outside options

Private cars and public transportation are two distinct and separate markets. Commuters who have access to alternative options outside public transit, such as taxis and private cars, are likely to have different mode choice preferences. We address this concern by rerunning the baseline models with a subsample of commuting trips by adult commuters who solely rely on public transport. More specifically, we first identify the round-trip journeys (rail or bus) between OD pairs made by adult commuters in a day, which include one round-trip journey made between 6 AM and 10 AM and another one between 4 PM and 8 PM. If a traveler repeats the round-trip journeys between the same OD pairs in at least 14 days during the month, we assume these journeys are made by commuters without outside options. The subsample of commuters without outside options makes up approximately 10% of the main sample. The results of the baseline model based on this subsample of commuters are reported in Table A-3 of the online Appendix.

The results are similar to those reported in Table 2, which implies that our results are robust and independent of the outside options of commuters. The coefficient on the difference in travel time unreliability for adult commuters is significant and negative, indicating that adult commuters on average prefer service reliability in their choice of public transport. For Group 1 adult commuters, the coefficient on the difference in travel time unreliability is negative and highly significant, while the coefficient on the difference in travel costs is negative and insignificant. The results suggest that Group 1 adult commuters care more about travel time unreliability than the average adult commuters, but travel costs are not the main consideration in their public transport mode decisions.[8] For Group 2 adult commuters, the coefficient on the difference in travel time unreliability is positive and marginally significant.

4.2.5 Submarket models

Singapore is a metropolis made up of multiple travel submarkets, as defined by different OD pairs. Our baseline model treats the island as a single integrated market, and the identification strategy does not account for heterogeneity in different submarkets. By focusing on selected submarkets, we ensure that commuters' public transport mode choices are not tainted by differences in the submarket's environment. We identify the submarkets in four popular locations in Singapore, which are Tampines, Bugis, Chinatown and Orchard Road, and track transit journeys based on origins or destinations in these four selected locations. Tampines is one of three regional centers in the eastern part of Singapore; Bugis is a popular shopping area in the civil district with many tourist attractions nearby; Chinatown is a historical district on the fringe of the CBD with several clusters of conservation shophouses; and Orchard Road is the main shopping belt, which is popular among tourists and local shoppers. We rerun the baseline models using the adult commuters in these four submarkets and report the results in Table A-4 of the online Appendix.

The submarket models generally produce similar patterns as those in the baseline models using the full samples. The coefficients on travel time and travel cost are negative and significant

[8] Since the subsample in this subsection consists of working adult commuters, the result for travel time unreliability is consistent with the results of subsequent tests, reported in Table 8, which predict that public transport commuters are sensitive to service reliability for work-related journeys.

in all four submarket models. Travel time unreliability is significant only in the Bugis model, and the coefficient has a negative sign.

4.2.6 Logit mode choice models

Logit models are widely used in travel mode choice modeling. We estimate the logit specifications of the baseline models as a robustness check. Commuters who choose only one single mode for all the journeys are dropped from the estimation due to constraints from the card fixed effects in the model. Therefore, the total sample size used in the logit models is smaller than that used in the corresponding baseline models. The top panel of Table 4 presents the results of the logit models using the standard deviation of travel time as the unreliability measure, and the bottom panel of Table 4 reports the results using the difference between the 50^{th} percentile and 90^{th} percentile of travel time as the unreliability measure.

As expected, the coefficients on travel time and travel cost are negative and significant in all the models. Travel time unreliability, as captured by different uncertainty measures, adversely affects the probability of choosing a given transit mode for the adult commuters, as well as for Group 1 adult computers in the two models. The difference between the 50^{th} percentile and 90^{th} percentile travel time has a negative but marginally significant (at the 0.1 level) impact on Group 2 adult commuters, indicating that they are only marginally sensitive to extreme delays. Students are only averse to travel time unreliability in the model using the standard deviation measure. The results for the senior commuter group are mixed. For the two different travel time unreliability measures, we find a positive and significant coefficient on the standard deviation of travel time at the 0.05 level, while the coefficient on the difference between the 50^{th} percentile and the 90^{th} percentile is negative and insignificant for the senior citizen group. The overall mode choice patterns in the logit models are similar to those in the linear probability models, which demonstrates that our findings are robust to different model specifications.

[Insert Table 4 here]

4.3 Value of travel time (VOT) and value of travel time reliability (VOR)

VOT measures commuters' willingness to pay (WTP) to shorten travel time, whereas VOR is the monetary value that commuters place on reducing travel time unreliability. Our estimates of WTP for travel time (or reliability) reflect commuters' trade-off between travel time (unreliability) and travel costs in their transport mode choices. Following the literature, we use Equations 6a and 6b to estimate the VOT and VOR across different commuter types based on the travel time coefficient (multiplied by 60 to convert the units from minutes to hours), the travel time unreliability coefficient and the travel cost coefficient estimated in the logit models with different travel time unreliability measures, as reported in Table 4. The VOT and VOR estimates are summarized in Table 5.

[Insert Table 5 here]

Based on the commuters' revealed preferences, the VOT for all the observations is estimated to be between S$1.615/hour and S$1.637/hour, on average, according to the logit models after controlling for the card fixed effects. Among the three types of commuters, adult commuters have the highest VOT, estimated to be between S$2.036/hour and S$2.081/hour, while the VOT estimates for senior citizens and children/students are much lower, between S$0.476/hour and S$0.533/hour and S$1.009/hour and S$1.100/hour, respectively. Among the adult commuters, the travel time savings are the highest, with savings between S$3.633 and S$4.580/hour for Group 1 compared to between S$1.724 and S$1.768/hour for Group 2. These results imply that the Group 1 frequent mode-switchers put the highest value on travel time among all adult commuters.

It is useful to express VOT and VOR as a fraction of the hourly wage rate. Official wage data for transit riders are not available in Singapore, and we instead use the Household Interview Travel Survey (HITS) data to estimate the median hourly wage of transit commuters. The HITS survey was conducted by the LTA in 2012, covering 1% of the households in Singapore, which includes over 10,000 households and 30,000 individuals. The survey data include detailed records on mobility, activity information, and the demographic characteristics of respondents who are above six years old.

We identify individuals who have used public transit at least once in the survey and use the information on their income level to calculate the median hourly wage by transit commuter type (adult, senior citizens, and child/student). The median hourly wage of all the adult transit commuters is computed to be approximately S$13.2, which is significantly lower than the average wage of S$22.3/hour for Singaporean residents.[9] The median hourly wages for senior citizens and children/students are reported to be zero. The VOT of adult transit commuters (S$2.0359-2.0806/hour) accounts for approximately 15.4% to 15.8% of the median hourly wage. The transit riders' VOT-to-wage rate ratio is lower than the ratio estimated for private car drivers in Western countries. The VOT-to-wage rate ratio for the average adult transit commuters in Singapore falls below the range of 20% to 100% of VOT for personal journeys, as documented in Small and Verhoef's (2007) study. However, if we use the median wage for Group 1 adult transit commuters, the VOT is estimated to be between 24.6% and 34.8% of the median hourly wage of the group. This result is comparable to Litman's (2008) results on the value of transit travel time, which varies between 25% and 35% of the prevailing wage rate under comfortable conditions. Factors such as demographics, cultural characteristics and quality of the transit system could explain the differences in the results.

The results of the logit models suggest that service reliability is statistically significant in predicting the mode choices of adult commuters. We affirm that the Group 1 adult commuter VOR is higher than the overall average VOR estimated for the sample. For travel time unreliability as

[9] In 2016, the median gross monthly income from work was S$4,056 and the weekly paid hours was 45.5. This gives a median hourly income of about S$22.3 (see http://www.singstat.gov.sg).

measured by the standard deviation of travel time, the VOR is estimated to be S$0.320/hour for all commuters, $S0.474/hour for the adult commuters, and significantly higher at S$4.065/hour for the Group 1 frequent mode-switcher adults.

The same pattern holds for the alternative unreliability measure based on the difference between the 50th and 90th percentiles of travel time, which are estimated at S$0.219/hour, S$0.286/hour, and S$1.868/hour for all commuters, adult commuters and Group 1 adult commuters, respectively. For adult commuters, the value of a reduction in service unreliability of one hour is equivalent to 2.1% and 3.6% of the median hourly wage, based on the standard deviation and the percentile range unreliability measures, respectively. For Group 1 adult commuters, the VOR fraction is higher at 30.8% of their median hourly wage for every one-hour reduction in unreliability as measured by the standard deviation of travel time, and 14.2% of their median hourly wage for every one-hour reduction in unreliability as measured by the differences in the 50th and 90th percentiles of travel time. The VOR estimates based on the standard deviation of travel time are higher than the estimates based on the difference between the 50th and 90th percentiles of travel time in the corresponding commuter groups. The differences in the VOR estimates could be attributed to cultural differences between Singapore and Western countries that are not related to commuter aversion to extreme delays.

The VOR-to-VOT ratio measures the relative importance of VOR over VOT. Group 1 adult commuters have a relatively higher VOR-to-VOT ratio than the overall commuter group and the adult commuter group. The VOR is computed to be 19.8% and 23.3% of the VOT for the overall commuter and the adult commuter groups, respectively. However, VOR is 1.12 times higher than VOT for the Group 1 adult frequent mode-switchers group. The same patterns are found in the models using the alternative reliability measure based on the difference between the 50th and 90th percentiles of travel time.

Conventional mode choice models could generate biased estimates if they fail to control for unobserved individual heterogeneity. For comparison purposes, we compute the VOT and VOR based on the logit models without the card fixed effects (bottom panel of Table 5)[10] and compare the results with those estimated from the models with the card fixed effects (top panel of Table 5). There are no significant differences in the overall behavioral patterns predicted by the two models; generally, we observe that the models without the card fixed effects produce higher VOT estimates relative to the estimates from the models with the card fixed effects. For Group 1 adult commuters, VOT is estimated at over S$6/hour for the models without the card fixed effects, which is compared to S$3.633/hour to S$4.580/hour for the models controlling for the card fixed effects. The VOR estimates are significantly different for the two models. Our results are consistent with the predictions of Brownstone and Small (2005) but contradict Hensher's (2001) comments that more sophisticated models produce more accurate VOT estimates, whereas simplified models underestimate VOT.

[10] The estimation results are not reported but are available upon request.

4.4 Elasticities

The elasticity of mode share with respect to the service level measures is defined as the percentage change in the bus mode share to a one-percent change in the service level. Based on the baseline model results in Table 2, we compute the elasticity of the bus mode share with respect to changes in travel times, travel time unreliability and travel fares by bus and rail. By Equation (7), the elasticity of the bus mode share change with respect to the bus travel time change is given by

$$\frac{\Delta(\text{bus share})/\text{bus share}}{\Delta(\text{bus travel time})/\text{bus travel time}} = \frac{\Delta(\text{bus share})}{\Delta(\text{bus travel time})}\frac{\text{bus travel time}}{\text{bus share}}$$
$$= \beta_1 \frac{\text{bus travel time}}{\text{bus share}},$$

where Δx is a small change in x.

The elasticity of the bus mode share change with respect to the bus travel time change for adult commuters is estimated to be approximately -1.46, and the bus mode share accounts for 0.110 of the journeys of adult commuters. If the bus travel time in each journey increases by one percent, the bus mode share of adult commuters' journeys falls by 1.46% from 0.110 to 0.108, i.e., [0.110 × (100%-1.46%)]. The results in Table 6 show significant heterogeneity in the elasticities across different commuter types. On average, the bus travel time elasticity for adult commuters (-1.46) is significantly higher than the elasticities for senior citizens (-0.70) and students/children (-0.79). Among the adult commuters, the travel time elasticity for Group 2 is higher than that for Group 1 because the bus mode share for Group 1 adult commuters is higher than that for Group 2 adult commuters.[11] The reliability elasticity for bus share is small for all commuter types except for Group 1 adult commuters (who have an elasticity of 0.22, where the opposite sign indicates the reliability level), who have a strong preference for certainty in work-related journeys (see Table 8). The bus fare elasticities for senior citizens and students/children are several times larger than the bus travel time elasticities. However, the bus travel time elasticities for senior citizens and students/children are lower than the corresponding elasticities for adult commuters, indicating that senior citizens and students/children are relatively more price sensitive than adult commuters.

[Insert Table 6 here]

Table 6 reports the cross-elasticity of the bus mode share change with respect to changes in rail service levels. The bus mode share elasticity with respect to the rail fare change is similar to the elasticity with respect to the bus fare change for all the commuter groups, whereas the elasticity with respect to rail travel time change is smaller than the elasticity with respect to bus travel time change. As implied by the elasticity equation and Figure 1 (ii), these results could be driven by the shorter rail travel times. The bus mode share change for Group 1 adult frequent mode-switchers is less sensitive to rail service unreliability but is more sensitive to bus service unreliability. As shown Figure 1 (iii), these results could be driven by the lower rail service unreliability.

[11] More specifically, by the results in Table 2, the coefficient on the travel time difference is most negative for Group 2 adults because they use rail for longer trips and save time this way and, by the results in Table 1, the denominator for Group 1 is a larger number in the above elasticity formula.

4.5 Normalized travel time unreliability

One possible explanation for the low estimates of the importance of travel time unreliability in subsection 4.1 could be the higher tolerance for travel time unreliability by commuters on longer trips. To test this hypothesis, we normalize the travel time unreliability variable by the travel distance and then recalibrate the baseline models.

The results are reported in Table 7. The results for the differences in travel time and travel costs are similar to the original baseline models. The coefficients on the two variables are negative and significant across all the commuter groups, except for Group 1 adult and student commuters, which are insensitive to travel costs. Unlike the early baseline models, which show that travel time unreliability as measured by the standard deviation of travel time is only significant for Group 1 adult commuters, the distance-normalized travel time unreliability coefficient is negative and significant in the full commuter sample model, implying that on average, commuters are sensitive to the normalized travel reliability measure. In the subgroup analyses, the normalized travel time unreliability of a given mode choice significantly reduces the probability of choosing that mode among the adults, including both Group 1 adults and Group 2 adults (at the 0.1 level), and the student commuters, whereas the impact on senior citizens is negative but insignificant.

[Insert Tables 7 here]

4.6 CBD effect

Commuters attach different opportunity costs (penalties) to different trip types, and their trip purpose influences their preferences for service reliability on their trips. As the purpose of each transit journey is not observable from the EZ-Link card transactions, we assume that journeys that either originate or end in the central business district (CBD) are work-related journeys that could inflict a "penalty" on commuters if they miss their schedules.

Next, we use the work-related journeys from/to the CBD to test the public transport mode choice for Group 1 adult commuters, controlling for the trips' origins and destinations. We include two dummy variables, "origin in CBD" and "destination in CBD", as well as their interaction terms with travel time unreliability as new explanatory variables in the model to test the impact of trip purpose on the preference for reliability. If the trip purpose is independent of the aversion to unreliability, the coefficients on the interaction terms are expected to be insignificant.

Columns (1), (3), (5), and (7) of Table 8 show that the results for travel time unreliability while controlling for the CBD dummy variables are similar to those in the models without the CBD dummy variables in subsection 4.1. The coefficient on travel time unreliability is insignificant and negative for all commuters, adult commuters and student commuters but is significant and positive for senior citizens. Columns (2), (4), (6), and (8) of Table 8 show the results of the model with the interaction terms for the difference in travel time unreliability and the CBD dummy variables. The coefficient on the original travel time unreliability variable captures the effect of service reliability for journeys with both the origin and the destination outside the CBD area, while the coefficients on the two interaction terms show the effects of service reliability for the CBD journeys (either originating from or ending in the CBD) relative to the non-CBD journeys. For adult commuters, the coefficients on the interaction terms are negative and significant, suggesting that travel time unreliability in a given mode significantly reduces the probability of choosing that mode for mostly work-related journeys starting from or ending in the CBD area relative to other journeys. However, the coefficients on the interaction terms for senior citizens and children/students, who are assumed to have no formal employment or work-related trips to the CBD, are mostly insignificant. These findings imply that preferences for reliability are correlated with trip purpose. Public transport riders are sensitive to service reliability for work-related journeys.

[Insert Table 8 here]

4.7 Rainfall shocks

The cross-sectional analyses of individual mode choice behavior and service-level preferences are susceptible to omitted variables and endogeneity problems. Commuters who choose to take a bus expect to have longer travel times because the bus will stop more frequently to allow boarding and alighting by commuters. To address these issues, we use rainfall at the origin bus stop as an exogenous shock and test the impact of expected bus travel time changes on the public transport mode choice of commuters.[12] Rainfall usually lasts for approximately 15 to 45 minutes in Singapore, and thunderstorms that come and dissipate quickly are hard to accurately forecast.[13] We use rainfall that is random and unpredictable as an exogenous shock in our tests.

We collected hourly rainfall data recorded at 61 weather stations in Singapore in August 2013. By dividing Singapore into equally sized cells of 200 m × 200 m each, we generated the hourly rainfall surface across Singapore using the Geographic Information Systems (GIS) tool. The interpolated rainfall value in each grid cell is an inverse-distance weighted average of the rainfall levels recorded at the six nearest weather stations. For example, Figure 2 plots the interpolated rainfall surface for the period from 1 pm to 2 pm on August 29, 2013. The rainfall value in the grid cells is matched to the bus stops/train stations at the origin location of a journey.

[Insert Figure 2 here]

We run two regressions on the observed travel time, one for all bus trips and another for all rail trips, to test the effects of rainfall on travel time. Given that rainfall at the origin location is likely to affect shorter bus trips, we create an interaction term between the rainfall level at the origin and travel distances of less than 1.6 km, which is the first quartile of all bus trips, and include this term as a regressor. We regress the log-travel time on the log-travel distance, the interaction term, origin-destination pair fixed effects, trip-starting hour fixed effects, and the interaction of the

[12] This experiment does not test if commuters switch the public transport mode during heavy-rain days, as we do not observe switching decisions in our data. We do not have data on the bus stop features and attributes at the origin location to identify if commuters are able to avoid unpleasant waiting at the bus stops. The experiment shows that rainy days could delay travel time and increase unreliability for OD trips by buses but not by rail for the same OD trip.

[13] To learn more about weather forecasting in Singapore, see http://www.nea.gov.sg/training-knowledge/weather-climate/weather-forecast/challenges-in-weather-forecasting

trip's starting hour and a weekday dummy. The results in Table A-5 of the online Appendix show that rainfall at the origin bus stop location increases travel time for short bus trips. A one mm/h (millimeter per hour) increase in rainfall leads to a 0.64% increase in bus travel time for a journey that is shorter than 1.6 km. For longer trips, the effect is insignificant because rain clouds in Singapore are typically zonal by nature, and rainfall occurrences could vary within a small geographical area across the island (see Figure 2). As expected, the impact of rainfall on rail travel time is insignificant.

To test the effects of rainfall at the origin of a journey on commuters' mode choice, we create an interaction term between a rainfall level dummy, which is defined by rainfall above a certain level, and a short-trip dummy, which represents a trip of less than 1.6 km. We add the interaction term to the public transport mode choice model of Eq. (7) and estimate the model using different rainfall levels to define the dummy and for all the adult commuters. We do not estimate models for other commuter groups due to the limited sample size; for instance, less than 0.5% of trips occur during a time with rainfall of 10 mm/hour or more. Column (2) of Table 9 shows that rainfall decreases the probability of choosing the bus mode in anticipation of possible increases in the bus travel time. However, Column (3) indicates that rainfall at the origin bus stop decreases the probability of choosing the bus mode only for short journeys. The rainfall effects on the bus mode choice increase with the rainfall levels. Column (4) shows that rainfall greater than 10 mm/h has relatively larger effects in reducing the probability of choosing the bus mode than rainfall less than 10 mm/h.

Using the models with rainfall of 10 mm/h or more as the reference, we estimate the elasticity measures (Table A-5 of the online Appendix) and show that rainfall causes the travel time to increase by 6.4% (i.e., 10 × 0.0064). We compute the bus mode share based on the elasticity of the bus share with respect to bus travel time, *e*, as estimated in subsection 4.4, as follows:

$$\begin{aligned}\Delta(\text{bus share})/\text{bus share} &= e \cdot \Delta(\text{bus travel time})/\text{bus travel time} \\ &= -1.4618 \cdot 0.064 = -0.0936.\end{aligned}$$

Given that the mean bus share for adults is 0.110 (see Table 1), this gives

$$\Delta(\text{bus share}) = -0.0936 \cdot \text{bus share} = -0.0936 \cdot 0.110 \approx -1\%.$$

The 1% drop in the bus share estimate is lower than the estimated fall in the bus share of 3.2% reported in Table 9, and the early result, though reasonable, may be an underestimation of the effect of bus travel time on mode choice.

[Insert Table 9 here]

5. Conclusion

We use a unique, large dataset containing transaction records for approximately four million smart transit cards in Singapore in August 2013 to study the travel preferences of public transit commuters. Due to the high-frequency nature of the dataset, we compute the VOT and VOR of public transit commuters, which cannot be done using traditional revealed-preference data. We find significant heterogeneity in preferences for travel time and reliability by types of public transit commuters and trip purposes. Our results are robust after controlling for unobserved individual heterogeneity, such as travel attitudes and lifestyle preferences, using the individual card fixed effects. We also use rainfall as an exogenous shock to test changes in commuters' behavioral responses to travel time uncertainties.

The results indicate that travel time and travel costs have a significant impact on the public transit mode choice of commuters. Travel time unreliability for a given transit mode significantly reduces the probability of that public transit mode being chosen by adults who are frequent switchers of transport modes. These adult frequent-switchers account for 21.4% of the adult transit commuters. However, travel time unreliability has an insignificant impact on other commuters. When using the distance-normalized travel time unreliability measure, we find that travel time has a significantly negative effect on the public transit mode choice for all commuters except for senior citizens. The preference for reliability is highly dependent on trip purpose. We find that travel time unreliability has stronger effects on mode choice for work-related trips, which are journeys with the origin and/or destination in the CBD, but the effect is insignificant for other trips.

Based on the logit model estimations, we compute the implied VOT and VOR for different transit commuter types. Adult commuters place a higher monetary value (approximately S$2/hour) on travel time savings than senior citizens (approximately S$0.5/hour) and students/children (approximately S$1/hour). Among the adult commuters, the frequent mode-switchers have the highest VOT, between S$3.63 and S$4.58/hour. On average, the VOT for adult transit commuters in Singapore relative to their median hourly wage is low, at approximately 15% to 16%, but this result is consistent with the belief that public transit commuters place a lower value on travel time savings than motorists.

There are two VOR measures for adult passengers: one is based on the standard deviation of travel time, and the other is based on the difference between the 50^{th} and 90^{th} percentiles of travel time, which are estimated at S$0.47/hour and S$0.29/hour, respectively. These numbers are translated into VOR-to-VOT ratios of 23.3% and 13.8%, respectively, which are close to the lower bound of the reliability ratio estimates, ranging from 10% to 251%, with a mean of 109%, as reported in previous studies (Carrion and Levinson, 2012). Among the adult passengers, the VOT, VOR and the reliability ratio (VOR/VOT) are higher for the frequent mode-switchers. The VOT of the frequent mode-switchers is estimated to be between S$3.63 and S$4.58/hour, which is approximately 24.6% to 34.8% of their median hourly wage. The VOR is estimated to be approximately 111.9% (based on the standard deviation of travel time) and 40.8% (based on the difference between the 50^{th} and 90^{th} percentiles of travel time) of their VOT.

The elasticities of mode share with respect to service attributes vary significantly across different commuter types. Adult commuters are more sensitive to travel time, while senior citizens and students/children are more sensitive to travel fare. The bus share elasticity with respect to bus travel time is estimated at -1.46 for adult commuters, -0.70 for senior citizens and -0.79 for students/children. The bus share elasticity with respect to the bus fare (transport cost) is estimated at -1.52 for adult commuters and -2.41 for senior citizens. A one-percent improvement in bus reliability increases the bus mode share of the adult frequent mode-switchers by 0.22%, but its impact on other commuters is insignificant.

Our estimates of the VOT and VOR for transit commuters are significantly lower than the previous estimates for motorists. Applying motorist-based VOT and VOR to public transport systems could misguide policy formulation; however, with the information on VOT and VOR based on transit commuters, transit authorities could design policies that encourage higher usage, reduce travel times and improve service reliability. The heterogeneity in public transit commuters' preferences demonstrates the needs for differentiated pricing schemes for different types of commuters and trip purposes in public transport systems. Having a "*one size fits all*" transport service may not induce optimal travel behaviors in public transport commuters, who have different willingness-to-pay for travel time and reliability. There is potential to introduce more customized services and cost-effective fare packages to reduce inefficiency in public transport service provisions and improve the service level satisfaction and utilization rate of public transport systems.

Singapore's government has experimented with various public transport policies targeted at shifting peak-load travel demand and improving efficiency in the public transport system. One such experiment is the early-bird free ride offered for MRT journeys into the CBD before 7:45 am on weekdays. While transit fare-based policies are more effective in changing the travel behaviors of cost-sensitive senior citizens and students/children, they are less effective in shifting the travel behaviors of adults, who are less sensitive to pre-peak hours incentives and discounts. Instead, improvements in travel time and service reliability could have a more direct impact on changing adult commuters' behaviors, especially for work-related journeys. For example, designated bus-only lanes during peak hours that reduce the travel time and improve the reliability of the bus mode could attract more adult commuters to switch to buses for work-related journeys. For the trains, more frequent trains to and from the CBD area would improve the travel time reliability and thus reduce disutility from transit services among disgruntled adult commuters.

References

Agarwal, S., Diao, M., Pan, J., Sing, T.F. 2015. Are Singaporean Cabdrivers Target Earners? *SSRN Working Paper*. [Available at https://ssrn.com/abstract=2338476].

Agarwal, Sumit and Charoenwong, Ben and Cheng, Shih-Fen and Keppo, Jussi, Fickle Fingers: Ride-Hail Surge Factors and Taxi Bookings (March 26, 2019). *SSRN Working Paper*. [Available at: https://ssrn.com/abstract=3157378].

Amin-Naseri, M.R., Baradaran, V. 2014. Accurate estimation of average waiting time in public transportation systems. *Transportation Services* 49(2): pp. 213-222.

Asensio, J., Matas, A. 2008. Commuters' valuation of travel time variability. *Transportation Research Part E: Logistics and Transportation Review* 44: pp. 1074-1085.

Bagchi, M., White, P.R. 2004. What role for smart-card data from bus system? Municipal Engineer, 157: pp. 39-46.

Bagchi, M., White, P.R. 2005. The potential of public transport smart card data. Transport Policy, 12: 464-474

Bates, J., Polak, J., Jones, P., Cook, A. 2001. The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review* 37: pp. 191-229.

Bhat, C.R., Sardesai, R. 2006. The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B: Methodological* 40: pp. 709-730.

Blythe, P. 2004. Improving public transport ticketing through smart cards. Proceedings of the Institute of Civil Engineers, Municipal Engineer, 157: 47-54

Brownstone, D., Small, K.A. 2005. Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice* 39: pp. 279-193.

Calfee, J., Winston, C. 1998. The value of automobile travel time: Implications for congestion policy. *Journal of Public Economics* 69: pp. 83-102.

Carrion, C., Levinson, D. 2012. Value of travel time reliability: a review of current evidence. *Transportation Research Part A: Policy and Practice* 46: pp. 720-741.

Diao, M., 2019. Towards sustainable urban transport in Singapore: Policy instruments and mobility trends, *Transport Policy*, 81, pp.320-330.

Diao, M., Leonard, D., Sing, T.F., 2017. Spatial-difference-in-differences models for impact of new Mass Rapid Transit Line on private housing values. *Regional Science and Urban Economics*, 67, pp.64-77.

Embrechts P., Kluppelberg C., Mikosch T. 1997. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

Huang, J., Levinson, D., Wang, J., Zhou, J. Wang, Z. 2018. Tracking job and housing dynamics with smartcard data. Proceedings of the National Academy of Sciences, 115(50): 12710-12715.

Keyhani, M.H., Schnee, M., Weihe, K. 2017. Arrive in time by train with high probability. *Transportation Science*.

Kitamura, R., Yamamoto, T., Susilo, Y.O., Axhausen, K.W. (2006). How routine is a routine? An analysis of day-to-day variability in prism vertex location. Transportation Research Part A, 40 (3): 259-279

Lam, T.C., Small, K.A. 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review* 37: pp. 231-251.

Lancester, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics* 95: 391-413.

Lee, S.G., Hickman, M., 2011. Travel pattern analysis using smart card data of regular users. Preprint CD-ROM for the 90th Annual Meeting of Transportation Research Board, Washington, DC.

Litman, T., 2008. Valuing transit service quality improvements. *Journal of Public Transportation*, 11 (2): 43-63.

Liu, H., Recker, W., Chen A. 2004. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A: Policy and Practice 38*: pp. 435-453.

Liu, H., He, X., Recker, W. 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. *Transportation Research Part B: Methodological* 41: pp. 448-461.

Mannering F.L., Hamed, M.M. 1990. Occurrence, frequency, and duration of commuters' work-to-home departure delay. *Transportation Research Part B: Methodological* 24(2): pp. 99-109.

Morency, C., Trépanier, M., Agard, B. 2007. Measuring transit use variability with smart-card data. Transport Policy, 14 (3): 193-203.

Noland, R.B., Small, K.A. 1995. Travel time uncertainty, departure time choice and the cost of morning commutes. *Transportation Research Record* 1493: pp.150-158.

Noland R.B, Small, K.A., Koskenoja, P., Chu, X.1998. Simulating travel reliability. *Regional Science and Urban Economics* 28(5): pp.535-564.

Pelletier, M.-P., Trepanier, M., Morency, C. 2011. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies.

Prakasam S. 2009. Evolution of e-payment in public transport – Singapore experience. *Journeys*, 53.

Shelfer, M., Procaccino, J.D. (2002). Smart card evolution. Communications of the ACM, 45 (7), pp. 83-88.

Small, K.A. 1982. The scheduling of consume activities: work trips. *American Economic Review* 72: pp. 467-479.

Small, K.A., Verhoef, E.T., 2007. The Economics of Urban Transportation. Routledge, London.

Small, K.A., Winston, C., Yan, J. 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica* 73(4): pp. 1367-1382.

Small, K.A., Winston, C., Yan, J. 2006. Differentiated road pricing, express lanes, and carpools: exploiting heterogeneous preferences in policy design. *Brookings-Wharton Papers on Urban Affairs* (7): pp. 53-96.

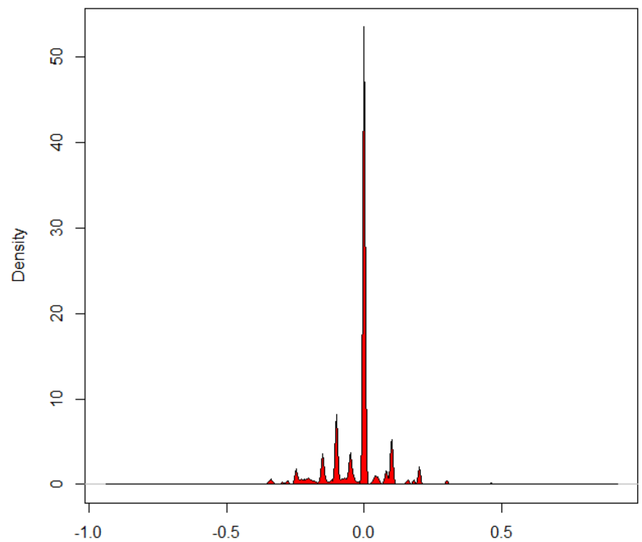
Small, K.A., Yan, J. 2001. The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation. Journal of Urban Economics, 49(2): pp. 310-336.

Song, S., Feng, C.-C., Diao, M., 2020. Vehicle quota control, transport infrastructure investment, and vehicle travel: A pseudo panel analysis. *Urban Studies*, 57(12), pp. 2527-2546.

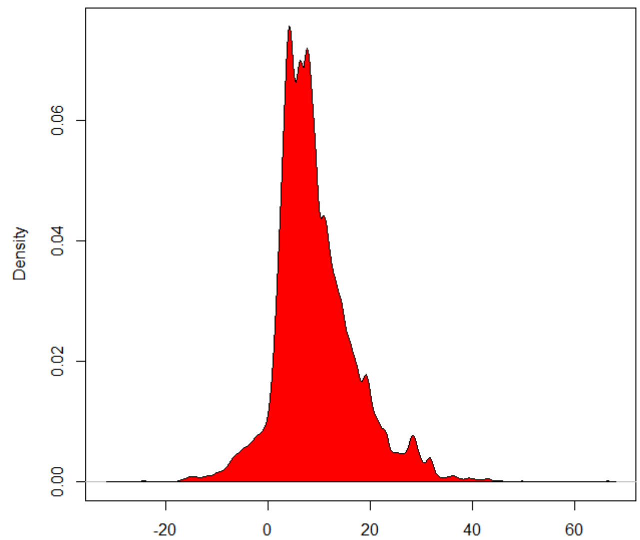
Yan, J. 2002. Heterogeneity in Motorists' Preferences for Travel Time and Time Reliability: Empirical Finding from Multiple Survey Data Sets and Its Policy Implications. Ph.D. Dissertation, University of California, Berkeley.

Yang, N., Lim, Y.L. 2017. Temporary Incentives Change Daily Routines: Evidence from a Field Experiment on Singapore's Subway. *Management Science*.

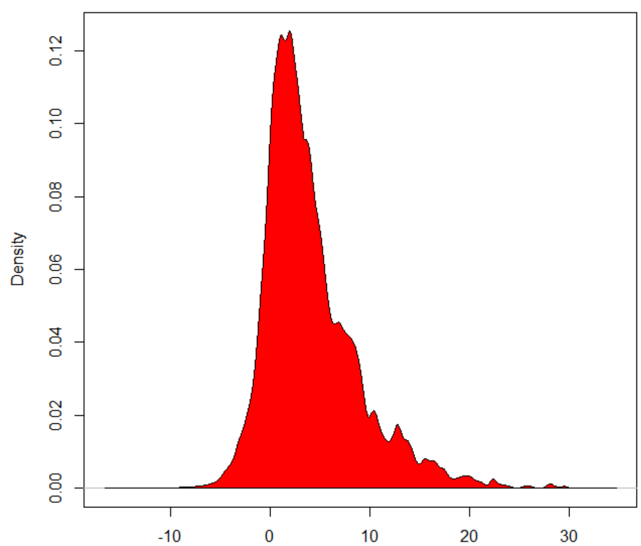
Zhu, Y., Diao, M., 2016. The impacts of urban Mass Rapid Transit lines on the density and mobility of high-income households: A Case Study of Singapore. *Transport Policy*, 51, pp.70-80.



(i) Difference in Travel Fares



(ii) Difference in Travel Times



(iii) Difference in Standard Deviations of Travel Times

Figure 1: Kernel Density for the Differences (Bus – MRT) in the Main Explanatory Variables.

Note: Bus travel time and standard deviation are in terms of minutes, bus and MRT fares are in terms of Singaporean dollars (SGD).

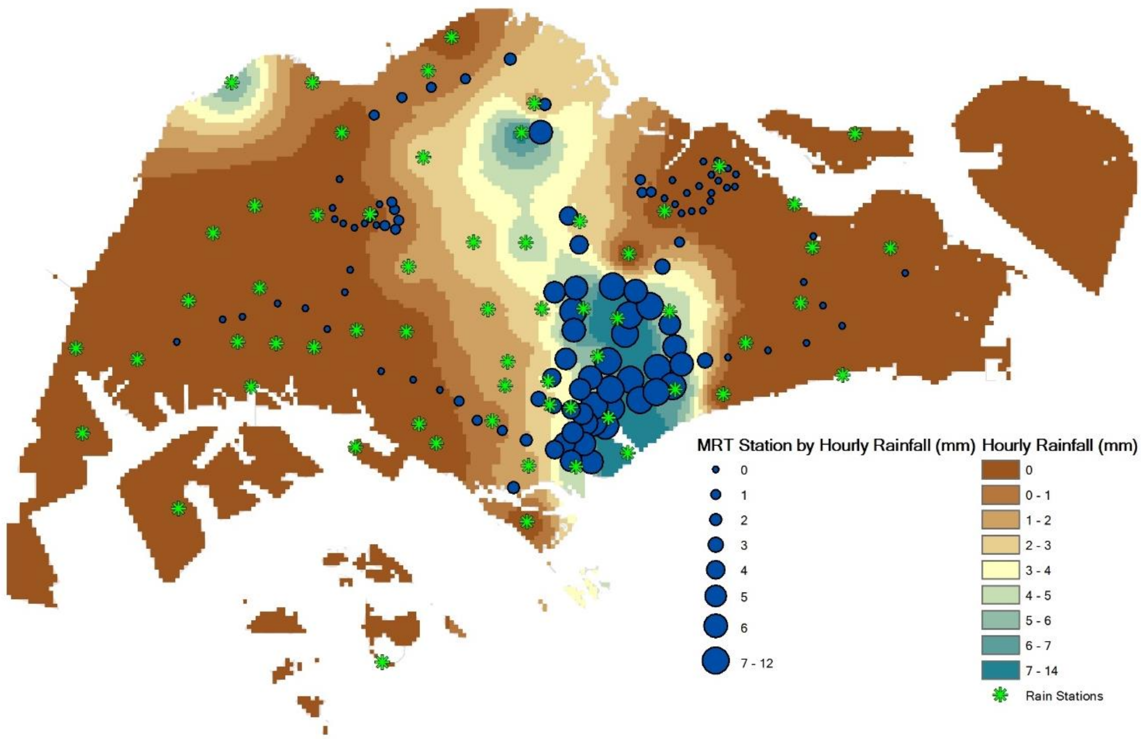


Figure 2: Interpolated Surface of Rainfall at 1-2pm on August 29, 2013.

The green stars are rain stations where actually rainfall levels are observed. The general color shows the interpolated levels at rain stations using geographic information system tools. The blue dots are rail stations. The size of a dot reflects the interpolated rainfall level at the station during 1-2 pm on August 29, 2013.

Table 1: Summary Statistics.

"Mode" is a binary variable, which equals 1 if the passenger chooses bus for the journey and 0 if he/she chooses rail. Δ(Travel time), Δ(Std. dev. of travel time), and Δ(Travel cost) are the differences in the expected travel times (bus travel time minus rail travel time), the travel time uncertainties (as measured by the standard deviations of the travel times), and the travel costs. "O_in_CBD" ("D_in_CBD") is a spatial dummy variable, which takes the value of 1 if the origin (destination) of the journey was in the central business district area and 0 otherwise. "Rainfall" is a continuous variable that measures the rainfall level at the boarding stop/station and the hour when the journey started. Adult passengers in Group 1 used multiple modes for at least one OD pair in the first 3 weeks of the month; Group 2 used multiple modes during the first 3 weeks of the month, but always stick to one single mode for an OD pair.

	All			Adult			Senior Citizen		
	Obs	Mean	Std Dev	Obs	Mean	Std Dev	Obs	Mean	Std Dev
Mode, bus=1 & rail=0	175,503	0.124	0.330	150,411	0.110	0.313	10,539	0.291	0.454
Δ(Travel time), minute	175,503	9.258	7.970	150,411	9.270	8.038	10,539	8.922	7.900
Δ(Std. dev. of travel time), minute	175,503	4.095	4.818	150,411	4.075	4.820	10,539	4.461	5.099
Δ(P90-P50 of travel time), min	175,503	2.750	8.893	150,411	2.743	8.937	10,539	2.816	9.480
Δ(Travel cost), SGD	175,503	-0.017	0.148	150,411	-0.019	0.158	10,539	-0.020	0.084
O_in_CBD, dummy	175,503	0.354	0.478	150,411	0.385	0.487	10,539	0.257	0.437
D_in_CBD, dummy	175,503	0.306	0.461	150,411	0.334	0.472	10,539	0.213	0.409
Rainfall, mm/h	175,503	0.130	1.340	150,411	0.129	1.343	10,539	0.161	1.471

	Student/Child			Adult: Group 1			Adult: Group 2		
	Obs	Mean	Std Dev	Obs	Mean	Std Dev	Obs	Mean	Std Dev
Mode, bus=1 & rail=0	14,553	0.151	0.358	7,228	0.261	0.439	11,683	0.204	0.403
Δ(Travel time), minute	14,553	9.378	7.284	7,228	7.795	7.424	11,683	8.298	8.214
Δ(Std. dev. of travel time), minute	14,553	4.038	4.581	7,228	3.642	4.368	11,683	3.850	4.648
Δ(P90-P50 of travel time), min	14,553	2.775	7.943	7,228	2.171	7.825	11,683	2.286	8.204
Δ(Travel cost), SGD	14,553	0.008	0.037	7,228	-0.023	0.195	11,683	-0.029	0.174
O_in_CBD, dummy	14,553	0.102	0.303	7,228	0.327	0.469	11,683	0.395	0.489
D_in_CBD, dummy	14,553	0.078	0.268	7,228	0.298	0.458	11,683	0.347	0.476
Rainfall, mm/h	14,553	0.123	1.205	7,228	0.182	1.664	11,683	0.186	1.669

Table 2: Baseline Models.

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in the standard deviations of the travel times, and the difference in the bus and rail travel costs. We estimate the model by OLS (see the discussion in the beginning of Section 4). The models are calibrated for all passengers (column 1), adults (column 2), senior citizens (column 5), and children/students (column 6), respectively. We further classify adults into three categories based on their mode choice behavior in the first 3 weeks of the month. Group 1 of adults switched modes for at least one OD pair during the 3 weeks. Group 2 of adults used both bus and rail during the 3 weeks, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the 3 weeks. We calibrate the same model for Groups 1 and 2 adults using their observations in the rest of the month and report the results in columns 3 and 4. Standard errors are reported in parentheses and are clustered at the card level. Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Name	**All**	**Adult**	**Adult: Group 1**	**Adult: Group 2**	**Senior**	**Student**
Δ(Travel time)	-0.0068***	-0.0068***	-0.0095***	-0.0100***	-0.0088***	-0.0056***
	(0.0003)	(0.0003)	(0.0018)	(0.0013)	(0.0018)	(0.0012)
Δ(Standard deviation of travel time)	-0.0003	-0.0004	-0.0072 ***	0.0000	0.0024	-0.0005
	(0.0003)	(0.0003)	(0.0021)	(0.0014)	(0.0018)	(0.0012)
Δ(Travel cost)	-0.2083***	-0.1906***	-0.2102	-0.3336***	-1.1151***	-0.2137
	(0.0244)	(0.0241)	(0.1996)	(0.0729)	(0.1967)	(0.1702)
Constant	0.1815***	0.1660***	0.3507***	0.2752***	0.3539***	0.2039***
	(0.0041)	(0.0043)	(0.0203)	(0.0155)	(0.0227)	(0.0139)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	175,503	150,411	7,228	11,683	10,539	14,553
R-squared	0.6101	0.5984	0.5880	0.7279	0.6773	0.5678
Adjusted R-squared	0.5282	0.5107	0.4243	0.5697	0.6287	0.4904

Table 3: Models with Alternative Reliability Measure

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in the difference between the 50th percentile and the 90th percentile of travel time, and the difference in the bus and rail travel costs. We estimate the model by OLS (see the discussion in the beginning of Section 4). The models are calibrated for all passengers (column 1), adults (column 2), senior citizens (column 5), and children/students (column 6), respectively. We further classify adults into three categories based on their mode choice behavior in the first 3 weeks of the month. Group 1 of adults switched modes for at least one OD pair during the 3 weeks. Group 2 of adults used both bus and rail during the 3 weeks, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the 3 weeks. We calibrate the same model for Groups 1 and 2 adults using their observations in the rest of the month and report the results in columns 3 and 4. Standard errors are reported in parentheses and are clustered at the card level. Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Name	**All**	**Adult**	**Adult: Group 1**	**Adult: Group 2**	**Senior**	**Student**
Δ(Travel time)	-0.0068***	-0.0068***	-0.0108***	-0.0098***	-0.0079***	-0.0059***
	(0.0003)	(0.0003)	(0.0017)	(0.0012)	(0.0016)	(0.0011)
Δ(P90-P50)	-0.0001	-0.0001	-0.0026***	-0.0004	0.0000	0.0003
	(0.0001)	(0.0001)	(0.0010)	(0.0006)	(0.0006)	(0.0005)
Δ(Travel cost)	-0.2083***	-0.1906***	-0.2115	-0.3330	-1.1101***	-0.2255
	(0.0245)	(0.0241)	(0.2014)	(0.0729)	(0.1960)	(0.1708)
Constant	0.1811***	0.1656***	0.3392***	0.2749***	0.3572***	0.2040***
	(0.0041)	(0.0043)	(0.0205)	(0.0155)	(0.0227)	(0.0141)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	175,503	150,411	7,228	11,683	10,539	14,553
R-squared	0.6101	0.5984	0.5868	0.7279	0.6770	0.5678
Adjusted R-squared	0.5282	0.5107	0.4227	0.5698	0.6284	0.4904

Table 4: Logit Models of Mode Choice

This table reports the results of logit specifications of the baseline models. The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in bus and rail travel time unreliability, and the difference in the bus and rail travel costs. We use two different travel time unreliability measures in the logit models: standard deviations of the travel times and the difference between the 50^{th} percentile and the 90^{th} percentile of travel time. The results are reported in Panels A and B, respectively. We include card and day fixed effects in all model. Passengers who only used one single mode for all the journeys are dropped in the estimation process. The models are calibrated for all passengers (column 1), adults (column 2), senior citizens (column 5), and children/students (column 6), respectively. We further classify adults into three categories based on their mode choice behavior in the first 3 weeks of the month. Group 1 of adults switched modes for at least one OD pair during the 3 weeks. Group 2 of adults used both bus and rail during the 3 weeks, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the 3 weeks. We calibrate the same model for Groups 1 and 2 adults using their observations in the rest of the month and report the results in columns 3 and 4. Standard errors are reported in parentheses. Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Name	**All**	**Adult**	**Adult: Group 1**	**Adult: Group 2**	**Senior**	**Student**
			Panel A			
Δ(Travel time)	-0.1042***	-0.1125***	-0.0698***	-0.1158***	-0.0949***	-0.0628***
	(0.0029)	(0.0033)	(0.0104)	(0.0108)	(0.0094)	(0.0081)
Δ(Standard deviation of travel time)	-0.0206***	-0.0262***	-0.0781***	-0.0101	0.0220**	-0.0235**
	(0.0037)	(0.0043)	(0.0140)	(0.0144)	(0.0111)	(0.0113)
Δ(Travel cost)	-3.8723***	-3.3158***	-1.1526***	-3.9296***	-10.6768***	-3.7370**
	(0.1807)	(0.1859)	(0.4289)	(0.6180)	(0.8578)	(1.6691)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	51,155	40,645	3,256	2,943	5,062	5,448
Log Likelihood	-15643.96	-12039.75	-1140.54	-869.60	-1614.12	-1884.16
Pseudo R-squared	0.1323	0.1523	0.0878	0.1979	0.1323	0.0420
			Panel B			
Δ(Travel time)	-0.1055***	-0.1149***	-0.0821***	-0.1133***	-0.0838***	-0.0691***
	(0.0027)	(0.0031)	(0.0100)	(0.0102)	(0.0088)	(0.0077)
Δ(P90-P50)	-0.0141***	-0.0158***	-0.0335***	-0.0162*	-0.0067	-0.0036
	(0.0022)	(0.0025)	(0.0084)	(0.0085)	(0.0062)	(0.0067)
Δ(Travel cost)	-3.8663***	-3.3130***	-1.0759**	-3.9457***	-10.5659***	-3.7675**
	(0.1808)	(0.1860)	(0.4283)	(0.6185)	(0.8535)	(1.6714)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	51,155	40,645	3,256	2,943	5,062	5,448
Log Likelihood	-15637.49	-12037.97	-1147.96	-867.93	-1615.49	-1886.20
Pseudo R-squared	0.1327	0.1524	0.0819	0.1994	0.1316	0.0410

Table 5: Value of Travel Time Saving and Reliability

The estimates are from equation (6) based on the estimation results of the logit model of mode choice. We only use the significant coefficients in Table 4. The values are calculated for all passengers, adults, senior citizens, and children/students, respectively. We further classify adults into three categories based on their mode choice behavior in the first 3 weeks of the month. Group 1 of adults switched modes for at least one OD pair during the 3 weeks. Group 2 of adults used both bus and rail during the 3 weeks, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the 3 weeks. The upper panel reports the estimates based on models with card fixed effect and the bottom panel reports the estimates based on models without card fixed effect.

Models with card fixed effects

	Standard Deviation			P90-P50		
	VOT (S$/hour)	VOR (S$/hour)	VOR/VOT	VOT (S$/hour)	VOR (S$/hour)	VOR/VOT
All	1.614	0.319	0.198	1.637	0.219	0.134
Adult	2.036	0.474	0.233	2.081	0.286	0.138
Adult: Group 1	3.633	4.065	1.119	4.580	1.868	0.408
Adult: Group 2	1.768			1.723		
Senior Citizen	0.533			0.476		
Student/Child	1.009			1.100		

Models without card fixed effects

	Standard Deviation			P90-P50		
	VOT (S$/hour)	VOR (S$/hour)	VOR/VOT	VOT (S$/hour)	VOR (S$/hour)	VOR/VOT
All	2.471	0.131	0.053	2.426	0.336	0.139
Adult	2.175	0.280	0.129	2.189	0.301	0.138
Adult: Group 1	6.052	2.773	0.458	6.273	0.940	0.150
Adult: Group 2	1.349			1.411		
Senior Citizen	0.942			0.874		
Student/Child	3.490			3.652		

Table 6: Bus Share Elasticities with Respect to Level of Service Measures.

The elasticity estimates are calculated from Table 2. * denotes insignificant coefficient for the corresponding level of service attribute in the base model in Table 2. Group 1 of adults switched modes for at least one OD pair during the first three weeks in the month. Group 2 of adults used both bus and rail during the first three weeks in the month, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the three weeks of the month.

Level of Service Measures	Adult	Adult: Group 1	Adult: Group 2	Senior Citizen	Student /Child
Bus travel time	-1.4618	-0.8087	-1.1628	-0.7006	-0.7853
Bus travel time unreliability	-0.0296	-0.2202	0.0004*	0.0713*	-0.0240*
Bus fare	-1.5182	-0.6927	-1.4402	-2.4144	-0.5864
MRT travel time	0.8912	0.5257	0.7566	0.4317	0.4394
MRT travel time unreliability	0.0152	0.1195	-0.0002*	-0.0349*	0.0116*
MRT fare	1.5507	0.7115	1.4883	2.4926	0.5752

Table 7: Models with Travel Time Unreliability Normalized by Travel Distance.

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in the standard deviations of the travel times normalized by travel distance, and the difference in the bus and rail travel costs. We estimate the model by OLS (see the discussion in the beginning of Section 4). The models are calibrated for all passengers (column 1), adults (column 2), senior citizens (column 5), and children/students (column 6), respectively. We further classify adults into three categories based on their mode choice behavior in the first 3 weeks of the month. Group 1 of adults switched modes for at least one OD pair during the 3 weeks. Group 2 of adults used both bus and rail during the 3 weeks, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the 3 weeks. We calibrate the same model for Groups 1 and 2 adults using their observations in the rest of the month and report the results in columns 3 and 4. Standard errors are reported in parentheses and are clustered at the card level. Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Name	**All**	**Adult**	**Adult: Group 1**	**Adult: Group 2**	**Senior**	**Student**
Δ(Travel time)	-0.0066***	-0.0066***	-0.0107***	-0.0098***	-0.0076***	-0.0051***
	(0.0003)	(0.0003)	(0.0016)	(0.0011)	(0.0016)	(0.0010)
Δ(Standard deviation of travel time normalized by distance)	-0.0037***	-0.0034***	-0.0147***	-0.0031*	-0.0033	-0.0061***
	(0.0004)	(0.0004)	(0.0027)	(0.0019)	(0.0021)	(0.0016)
Δ(Travel cost)	-0.2049***	-0.1875***	-0.1923	-0.3313***	-1.0977***	-0.2075
	(0.0242)	(0.0239)	(0.1941)	(0.0724)	(0.1963)	(0.1687)
Constant	0.1857***	0.1698***	0.3636***	0.2796***	0.3619***	0.2107***
	(0.0041)	(0.0043)	(0.0202)	(0.0157)	(0.0230)	(0.0139)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	175,503	150,411	7,228	11,683	10,539	14,553
R-squared	0.6108	0.5990	0.5909	0.7282	0.6773	0.5695
Adjusted R-squared	0.529	0.5115	0.4283	0.5701	0.6288	0.4924

Table 8: Models with CBD Effect.

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the travel times (bus travel time minus rail travel time), the difference in the standard deviations of the travel times, the difference in the travel costs, origin/destination in CBD, and interaction terms of origin/destination in CBD and the difference in the standard deviations. The models are calibrated for all passengers, adults, senior citizens, and children/students by OLS (see the discussion in the beginning of Section 4). Standard errors are reported in parentheses and are clustered at the card level. Significance code: *** 0.01; ** 0.05; * 0.1.

	All		Adult		Senior Citizen		Student/Child	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ(Travel time)	-0.0073***	-0.0072***	-0.0072***	-0.0072***	-0.0090***	-0.0086***	-0.0061***	-0.0060***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0007)	(0.0007)	(0.0006)	(0.0006)
Δ(Standard deviation of travel time)	0.0001	0.0016***	-0.0001	0.0016***	0.0024***	0.0050***	-0.0001	0.0003
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0009)	(0.0010)	(0.0007)	(0.0007)
Δ(Travel cost)	-0.1953***	-0.1984***	-0.1786***	-0.1819***	-1.0177***	-1.0371***	-0.2351**	-0.2221**
	(0.0070)	(0.0071)	(0.0069)	(0.0069)	(0.0622)	(0.0624)	(0.1038)	(0.1042)
Origin in CBD	0.0086***	0.0211***	0.0040**	0.0155***	0.0881***	0.1359***	0.0036	0.0055
	(0.0019)	(0.0023)	(0.0019)	(0.0024)	(0.0104)	(0.0128)	(0.0115)	(0.0138)
Destination in CBD	-0.0240***	-0.0190***	-0.0238***	-0.0187***	0.0027	-0.0021	-0.0609***	-0.0452***
	(0.0020)	(0.0024)	(0.0020)	(0.0024)	(0.0110)	(0.0138)	(0.0129)	(0.0158)
Origin in CBD * Travel time unreliability		-0.0031***		-0.0028***		-0.0103***		-0.0007
		(0.0003)		(0.0003)		(0.0016)		(0.0021)
Destination in CBD * Travel time unreliability		-0.0012***		-0.0013***		0.0006		-0.0039*
		(0.0003)		(0.0003)		(0.0016)		(0.0023)
Constant	0.1890***	0.1819***	0.1759***	0.1688***	0.3345***	0.3201***	0.2104***	0.2078***
	(0.0036)	(0.0036)	(0.0037)	(0.0038)	(0.0173)	0.0175	(0.0132)	(0.0132)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	175,503	175,503	150,411	150,411	10,539	10,539	14,553	14,553
R-squared	0.6105	0.6109	0.5989	0.5993	0.6799	0.6813	0.5686	0.5688
Adjusted R-squared	0.5287	0.5292	0.5113	0.5118	0.6317	0.6333	0.4912	0.4913

Table 9: Rainfall and Mode Choice of Adult Passengers.

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in travel times (bus travel time minus rail travel time), the difference in standard deviations of the travel times, and the difference in the travel costs. “Rain” is the rainfall level in millimeter in the hour of boarding. “Short_trip” is a dummy variable indicating that the trip distance is less than 1.6km. The models are calibrated for adult passengers. Standard errors are reported in parentheses and are clustered at the card level Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)		(2)		(3)		(4)	
Δ(Travel time)	-0.0068	***	-0.0068	***	-0.0068	***	-0.0068	***
	(0.0003)		(0.0003)		(0.0003)		(0.0003)	
Δ(Standard deviation of travel time)	-0.0004		-0.0004		-0.0004		-0.0004	
	(0.0003)		(0.0003)		(0.0003)		(0.0003)	
Δ(Travel cost)	-0.1906	***	-0.1904	***	-0.1902	***	-0.1902	***
	(0.0241)		(0.0241)		(0.0241)		(0.0241)	
Rain>0			-0.0144	***	-0.0056		-0.0056	
			(0.0050)		(0.0063)		(0.0063)	
(Rain>0) * Short_trip					-0.0243	**		
					(0.0106)			
(0<Rain<10) * Short_trip							-0.0218	*
							(0.0121)	
(Rain>10) * Short_trip							-0.0321	**
							(0.0147)	
Constant	0.1660	***	0.1660	***	0.1661	***	0.1661	***
	(0.0043)		(0.0043)		(0.0043)		(0.0043)	
Card fixed effect	Yes		Yes		Yes		Yes	
Day fixed effect	Yes		Yes		Yes		Yes	
Num. of observations	150,411		150,411		150,411		150,411	
R-squared	0.5984		0.5984		0.5985		0.5985	
Adjusted R-squared	0.5107		0.5108		0.5108		0.5108	

Online Appendix

Preferences of Public Transit Commuters: Evidence from Smart Card Data in Singapore

Sumit Agarwal[1], Mi Diao[2], Jussi Keppo[3], Tien Foo Sing[4]

16 May 2020

Abstract

This study employs an administrative dataset containing high-frequency transaction records of approximately four million smart transit cards of Singaporean residents to study travel preferences of public transport commuters. We examine impact of service attributes including travel time, reliability and travel cost on commuters' transport mode choices and estimate the implied value of travel time (VOT) and value of reliability (VOR) for different types of public transit commuters. The results show significant heterogeneity in transport preference of commuters with adult transit commuters having higher VOT and VOR relative to senior citizens, students and children. Among the adult group, commuters who frequently switch their modes have significantly higher VOT and VOR. The results have important implications for policymakers in formulating strategies to improve efficiency of the public transport system; and having flexible and customized transport services could increase utility and satisfaction levels of public transit commuters.

Keywords: *Public transportation, travel preference, value of travel time, value of reliability, big data*

---

1 Department of Finance, Business School, National University of Singapore. Email:ushakri@yahoo.com
2 College of Architecture and Urban Planning, Tongji University, China. Email: diaomi@gmail.com
3 Business School, Institute of Operations Research and Analytics, and Institute of Data Science, National University of Singapore. Email: keppo@nus.edu.sg
4 Department of Real Estate and Institute of Real Estate and Urban Studies (IREUS), National University of Singapore. Email: rststf@nus.edu.sg

We would like to thank the Land Transport Authority (LTA) for sharing the main empirical data, and Singapore-MIT Alliance in Research and Technology (SMART) for additional data support.

This online Appendix contains additional robustness results that are not reported in the main paper due to the space constraints. The additional results supplement the main findings of the paper, and they are summarized in the Tables below:

Table A-1: Models with Alternative Cutoff Distances
Table A-2: Models for Infrequent Transit Riders
Table A-3: Mode Choice of Adult Passengers without Outside Option
Table A-4: Submarket Models
Table A-5: Rainfall Level and Travel Time

A1. Different cut-off distances from rail stations

To check the robustness of the cut-off distance assumption, we use different cut-off distances of 150 m, 180 m and 200 m to rail stations to filter our data and rerun the models, respectively. The results are reported in Table A1.

The main findings still hold in the models using the alternative cut-off distances. However, we observe differences in the results between the alternative models and the baseline models. In the baseline model, travel time unreliability is only significant in Group 1 adult commuter model. As we reduce the buffer size from the rail stations, commuter in different subgroups become more sensitive to the unreliability measure.

In the models with the smallest buffer size of 150 m, the adult group shows significant aversion to travel time unreliability at the 0.05 level. The child/student group has a negative coefficient that is marginally significant at the 0.1 level. For Group 1 adult commuters, the negative effects of the estimated travel time unreliability on the mode choice probability generally increase with the cut-off distance to rail stations. A 10-minute increase in the standard deviation of travel time reduces the probability of mode choice by 5.5, 5.3, 6.2 and 7.2 percent when the cut-off distance increase from 150 m to 180 m, 20 0m, and 215 m, respectively.

A2. Mode choice of infrequent transit commuters

Our baseline models are estimated based on a sample of frequent transit commuters who have made at least 30 transit journeys during the month. These commuters are less likely to have outside options (such as private cars, taxis, etc.) compared to other infrequent transit commuters. To compare differences in preferences of service attributes by commuters having different transit trip frequencies, we run the baseline models using the sample of infrequent transit commuters and report the results in Table A2.

We find that the mode choice behaviors of infrequent transit commuters are not significantly different from those of frequent commuters. The coefficients on the travel time variable are significant and negative in all the subgroups implying that travel time generates disutility to all infrequent transit commuters. Travel costs reduce the probability of the transit mode choice among all the infrequent transit commuters, except for children and students. The impact of travel time unreliability is negative but insignificant for all the subgroups of infrequent transit commuters.

A3. Mode choice of transit commuters without outside options

Private cars and public transportation are two distinct and separate markets. Commuters who have access to alternative options outside the public transit, such as taxi and private car, are likely to have different mode choice preferences. We address this concern by re-running the baseline models with a subsample of commuting trips by adult commuters who solely rely on public transport. More specifically, we first identify the round-trip journeys (rail or bus) by the OD pairs made by adult commuters in a day, which include one round-trip journey made between 6 am and 10 am and another one between 4 pm and 8 pm. If a traveler repeated the round-trip journeys by the same OD pairs in at least 14 days during the month, we assume these journeys are made by commuters without outside options. The subsample of commuters without outside options makes up about 10% of the main sample. The results of the baseline model based on this subsample of commuters are reported in Table A3.

The results imply that our results are robust and independent of the outside options of commuters. The coefficient on the travel time unreliability difference for adult commuters is significant and negative indicating that adult commuters on average prefer service reliability in their choice of public transport. For Group 1 adult commuters, the coefficient on travel time unreliability difference is negative and highly significant, while the coefficient on travel cost difference is negative and insignificant. The results suggest that Group 1 adult commuters care more about travel time unreliability than the average adult commuters, but travel costs are not the main consideration in their public transport mode decisions. For Group 2 adult commuters, the coefficient on travel time unreliability difference is positive and marginally significant.

A4. Submarket models

Singapore is a metropolitan made up of multiple travel submarkets as defined by different OD pairs. Our baseline model treats the island as a single integrated market, where the identification strategy does not account for heterogeneity in different submarkets. By focusing on selected submarkets, we ensure that commuters' public transport mode choices are not tainted by differences in the submarket's environment. We identify the submarkets in four popular locations in Singapore, which are Tampines, Bugis, Chinatown and Orchard Road, and track the transit journeys based on origins or destinations in these four selected locations. Tampines is one of the three regional centers in the eastern part of Singapore; Bugis is a popular shopping area in the civil district with many tourist attractions nearby; Chinatown is a historical district at the fringe of the CBD with several clusters of conservation shophouses; and Orchard Road is the main shopping belt that is popular among tourists and local shoppers. We rerun the baseline models using the adult commuters in these four submarkets and report the results in Table A4.

The submarkets models generally produce similar patterns as in the early baseline models using the full samples. The coefficients on travel time and travel cost are negative and significant in all the four submarkets models. The travel time unreliability is significant only in the Bugis model, and the coefficient has a negative sign.

A5. Rainfall Shocks

We run two regressions of the observed travel time, one for all bus trips and another one for all rail trips separately, to test the effects of rainfall on the travel time. Given that rainfalls at the origin location are likely to affect shorter bus trips, we create an interaction term of the rainfall level at the origin and the travel distance of less than 1.6 km, which is the first quartile of all bus trips, and include the term as a regressor. We regress the log-travel time on the log-travel distance, the interaction term, the origin-destination pair fixed effects, the trip-starting hour fixed effects, and the interaction of trip's starting hour and a weekday dummy. The results in Table A.5 show that rainfalls at the origin bus stop location increase the travel time for short bus trips. One mm/h (millimeter per hour) increase in rainfall leads to 0.64% increase in bus travel time for a journey that is shorter than 1.6 km. For longer trips, the effect is insignificant because rain clouds in Singapore are typically zonal by nature, and rainfall occurrences could vary within a small geographical area across the island (see Figure 2). As expected, the impact of rainfalls on the rail travel time is insignificant.

Table A-1: Models with Alternative Cutoff Distances

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in the standard deviations of the travel times, and the difference in the bus and rail travel costs. We estimate the model by OLS (see the discussion in the beginning of Section 4). We use 3 different cut-off distances, 150m, 180m, and 200m to exclude bus journeys faraway from MRT stations. The results are reported in panels A, B, and C, respectively. The models are calibrated for all passengers (column 1), adults (column 2), senior citizens (column 5), and children/students (column 6), respectively. We further classify adults into three categories based on their mode choice behavior in the first 3 weeks of the month. Group 1 of adults switched modes for at least one OD pair during the 3 weeks. Group 2 of adults used both bus and rail during the 3 weeks, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the 3 weeks. We calibrate the same model for Groups 1 and 2 adults using their observations in the rest of the month and report the results in columns 3 and 4. Standard errors are reported in parentheses and are clustered at the card level. Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Name	**All**	**Adult**	**Adult: Group 1**	**Adult: Group 2**	**Senior**	**Student**
			Panel A: Cut-off = 150m			
Δ(Travel time)	-0.0039***	-0.0039***	-0.0065***	-0.0056***	-0.0064***	-0.0021*
	(0.0002)	(0.0002)	(0.0017)	(0.0011)	(0.0019)	(0.0011)
Δ(Standard deviation of travel time)	-0.0006**	-0.0005**	-0.0055***	-0.0006	0.0011	-0.0019*
	(0.0002)	(0.0002)	(0.0020)	(0.0011)	(0.0017)	(0.0011)
Δ(Travel cost)	-0.1436***	-0.1308***	-0.0697	-0.2514***	-0.9403***	-0.2566*
	(0.0221)	(0.0219)	(0.1791)	(0.0743)	(0.2176)	(0.1405)
Constant	0.1089***	0.1008***	0.2261***	0.1493***	0.2105***	0.1182***
	(0.0034)	(0.0036)	(0.0183)	(0.0139)	(0.0206)	(0.0125)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	165,605	143,014	6,397	10,543	9,003	13,588
R-squared	0.5907	0.5818	0.5829	0.7321	0.6765	0.5388
Adjusted R-squared	0.5005	0.4866	0.4003	0.5638	0.6219	0.4510
			Panel B: Cut-off = 180m			
Δ(Travel time)	-0.0047***	-0.0047***	-0.0071***	-0.0065***	-0.0075***	-0.0031***
	(0.0003)	(0.0003)	(0.0018)	(0.0012)	(0.0019)	0.0012
Δ(Standard deviation of travel time)	-0.0005*	-0.0004*	-0.0053***	-0.0012	0.0013	-0.0013
	(0.0003)	(0.0003)	(0.0020)	(0.0013)	(0.0018)	(0.0011)
Δ(Travel cost)	-0.1671***	-0.1516***	-0.1428	-0.2781***	-1.0378***	-0.2303
	(0.0225)	(0.0222)	(0.1851)	(0.0723)	(0.2075)	(0.1545)
Constant	0.1314***	0.1202***	0.2640***	0.1866***	0.2704***	0.1464***
	(0.0037)	(0.0038)	(0.0194)	(0.0142)	(0.0210)	(0.0137)

Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	168,632	145,177	6,661	10,892	9,538	13,917
R-squared	0.5959	0.5854	0.5792	0.7313	0.6779	0.5415
Adjusted R-squared	0.5081	0.4923	0.4012	0.5667	0.6254	0.4560
			Panel C: Cut-off = 200m			
Δ(Travel time)	-0.0060***	-0.0059***	-0.0094***	-0.0082***	-0.0092***	-0.0044***
	(0.0003)	(0.0003)	(0.0018)	(0.0012)	(0.0018)	(0.0012)
Δ(Standard deviation of travel time)	-0.0003	-0.0004	-0.0062***	-0.0009	0.0024	-0.0008
	(0.0003)	(0.0003)	(0.0020)	(0.0014)	(0.0018)	(0.0011)
Δ(Travel cost)	-0.2153***	-0.1972***	-0.2103	-0.3439***	-1.1544***	-0.3678**
	(0.0238)	(0.0234)	(0.2011)	(0.0737)	(0.1982)	(0.1702)
Constant	0.1580***	0.1448***	0.3253***	0.2369***	0.3092***	0.1747***
	(0.0039)	(0.0041)	(0.0200)	(0.0151)	(0.0213)	(0.0137)
Card fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Num. of observations	172,298	147,985	6,974	11,341	10,095	14,218
R-squared	0.6084	0.5959	0.5883	0.7315	0.6908	0.5608
Adjusted R-squared	0.5249	0.5066	0.4214	0.5722	0.6426	0.4805

Table A-2: Models for Infrequent Transit Riders

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in the standard deviations of the travel times, and the difference in the bus and rail travel costs. We estimate the model by OLS (see the discussion in the beginning of Section 4) using a 10% random sample of all infrequent transit riders that had less than 30 transit journeys in the month. The models are calibrated for all passengers (column 1), adults (column 2), senior citizens (column 3), and children/students (column 4), respectively. Standard errors are reported in parentheses and are clustered at the card level. Significance code: *** 0.01; ** 0.05; * 0.1.

	All infrequent transit riders (<30 public transport journeys in the month)							
	(1)		(2)		(3)		(4)	
Name	**All**		**Adult**		**Senior**		**Student**	
Δ(Travel time)	-0.0070	***	-0.0071	***	-0.0064	***	-0.0068	***
	(0.0003)		(0.0003)		(0.0011)		(0.0008)	
Δ(Standard deviation of travel time)	-0.0003		-0.0003		-0.0007		-0.0002	
	(0.0002)		(0.0003)		(0.0011)		(0.0009)	
Δ(Travel cost)	-0.2997	***	-0.2771	***	-1.0520	***	-0.1115	
	(0.0228)		(0.0232)		(0.1107)		(0.1397)	
Constant	0.2363	***	0.2106	***	0.4631	***	0.2835	***
	(0.0051)		(0.0056)		(0.0212)		(0.0163)	
Card fixed effect	Yes		Yes		Yes		Yes	
Day fixed effect	Yes		Yes		Yes		Yes	
Num. of observations	337,667		280,154		22,779		34,734	
R-squared	0.7554		0.7438		0.8058		0.7188	
Adjusted R-squared	0.5647		0.5374		0.6724		0.5359	

Table A-3: Mode Choice of Adult Passengers without Outside Option

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in the standard deviations of the travel times, and the difference in the bus and rail travel costs. We estimate the model by OLS. The models are calibrated for adults (column 1), Group 1 of adults (column 2), and Group 2 of adults (3), respectively. Group 1 of adults switched modes for at least one OD pair during the month. Group 2 of adults used both bus and rail during the month, but stick to a particular mode (bus or rail) in each OD pair. Group 3 of adults used one single mode (bus or rail) for all the journeys in the month. Standard errors are reported in parentheses and clustered at the day level. Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)	(2)	(3)
Name	**Adult**	**Adult: Group 1**	**Adult: Group 2**
Δ(Travel time)	-0.0034***	-0.0086***	-0.0098***
	(0.0006)	(0.0030)	(0.0012)
Δ(Standard deviation of travel time)	-0.0011**	-0.0101***	0.0021**
	(0.0004)	(0.0026)	(0.0010)
Δ(Travel cost)	-0.0929***	-0.1500	-0.1996***
	(0.0134)	(0.0914)	(0.0161)
Card fixed effect	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes
Num. of observations	21,488	3,076	2,729
R-squared	0.8459	0.6923	0.9713
Adjusted R-squared	0.8299	0.6585	0.9678

Table A-4: Submarket Models

The dependent variable is a dummy variable, which takes a value of 1 if taking bus, and 0 if taking rail. The independent variables are the difference in the bus and rail travel times (bus travel time minus rail travel time), the difference in the standard deviations of the travel times, and the difference in the bus and rail travel costs. We estimate the model by OLS (see the discussion in the beginning of Section 4). The models are calibrated for journeys by adult passengers that started or ended at four locations, Tampines, Bugis, Chinatown and Orchard Road, respectively. Standard errors are reported in parentheses and are clustered at the card level. Significance code: *** 0.01; ** 0.05; * 0.1.

	(1)		(2)		(3)		(4)	
	Tampines		**Bugis**		**Chinatown**		**Orchard Road**	
Name	**Adult**		**Adult**		**Adult**		**Adult**	
Δ(Travel time)	-0.0074	***	-0.0053	***	-0.0108	***	-0.0030	***
	(0.0014)		(0.0007)		(0.0013)		(0.0009)	
Δ(Standard deviation of travel time)	0.0011		-0.0023	**	-0.0018		0.0002	
	(0.0011)		(0.0010)		(0.0013)		(0.0008)	
Δ(Travel cost)	-0.2374	***	-0.7549	***	-0.3075	***	-0.2336	***
	(0.0858)		(0.0935)		(0.0893)		(0.0818)	
Constant	0.2122	***	0.1547	***	0.2684	***	0.1225	***
	(0.0250)		(0.0145)		(0.0207)		(0.0144)	
Card fixed effect	Yes		Yes		Yes		Yes	
Day fixed effect	Yes		Yes		Yes		Yes	
Num. of observations	7,978		22,013		13,741		20,786	
R-squared	0.8358		0.7827		0.8133		0.7485	
Adjusted R-squared	0.7766		0.6564		0.6989		0.5753	

Table A-5: Rainfall Level and Travel Time

The dependent variable is the log of travel time by bus or rail. Log(distance) is log of travel distance (km), Rainfall is rainfall amount at the origin (mm/h), and Short_Distance is a dummy variable that takes value of 1 if the trip distance is less than 1.6km. We control for OD fixed effect, hour of day fixed effect, and the interaction term of hour of day and a weekday dummy. We run two regressions of the observed travel time, one for all bus trips and another one for all rail trips separately. Standard errors are reported in parentheses. Significance code: *** 0.01; ** 0.05; * 0.1.

	Bus		**Rail**	
Log(distance)	1.1718	***	-0.1449	***
	(0.0078)		(0.0023)	
Rainfall * (Short_Distance)	0.0064	***	0.0014	
	(0.0023)		(0.0023)	
Fixed effect				
Origin-Destination	Yes		Yes	
Hour of day	Yes		Yes	
(Hour of day) * weekday	Yes		Yes	
Number of observations	20,827		176,438	
R squared	0.939		0.862	