Transportation
https://doi.org/10.1007/s11116-020-10089-w

Longitudinal analysis of activity generation in the Greater Toronto and Hamilton Area

Gozde Ozonder[1] · **Eric J. Miller**[1]

Abstract
This paper presents a longitudinal analysis of activity generation behaviour in the Greater Toronto and Hamilton Area (GTHA) between 1996 and 2016 for various activity types: work, school, shopping, other. The analyses are conducted using the data from the five most recent Transportation Tomorrow Surveys. For work and school purposes, the population is divided into sub-categories considering occupational sectors and educational levels respectively. Further subdivision is made by treating first work/school activity of the day and subsequent work/school activities as distinct activity types. Considerable stability over time in the majority of the model parameters is found in all cases, indicating that both work/school and non-work/school activity episode generation in the GTHA has been very stable over the 20-year period analyzed. Year-specific models and joint models, within which the data are pooled across the years, return very similar results implying that robust joint models that exploit the full time-series of survey data available can be constructed. While first-trips to work and post-secondary schools in the day can be parametrically modelled with reasonable fits, second/subsequent work/school activities and non-work/school activities display considerable randomness in occurrence. Elementary and secondary school trips generally need only be modelled using average trip rates across the student population: parametric, utility-based models provide very little additional explanatory power. In addition, investigation of survey design biases shows that there is no significant survey design effect on activity/trip generation for the first work/school-related activities, however, the models reveal significant biases when the subsequent work/school-related activities and non-work/school activities are analyzed.

Keywords Activity generation · Longitudinal analysis · Work/school activities · Shopping · Survey design bias · GTHA

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11116-020-10089-w) contains supplementary material, which is available to authorized users.

✉ Gozde Ozonder
gozde.ozonder@mail.utoronto.ca

Eric J. Miller
miller@ecf.utoronto.ca

[1] Department of Civil and Mineral Engineering, Universiy of Toronto, Toronto M5S 1A4, Canada

Published online: 24 February 2020

Introduction

Activity-based travel demand models are increasingly used to evaluate land-use, transportation and environment related policy and investment decisions, gradually replacing traditional four-step models in many urban regions (Castiglione et al. 2015). Major advantages of these models over four-step models are their capability to incorporate time and space constraints, and to link travel and activity behaviour of individuals within a household considering the interdependencies between the household members. A common feature of all activity-based models is an activity generation component (Miller 2018), which captures the needs to engage in different types of activities (Habib 2007). The conventional four-step model equivalent of this component is trip generation, the first stage of the demand modelling process, where the number of trips produced by and attracted to each geographical zone is estimated (Castiglione et al. 2015).

As the determinant of overall trip levels, activity/trip generation is of critical importance in all demand forecasting applications. Typically, all components of a travel demand model, including activity generation, are developed using a single, possibly the most recent (Mwakalonge and Badoe 2012), cross-sectional data set (Hu 2010) with the assumption that the model parameters will remain stable for forecasting applications (de Dios Ortúzar and Willumsen 2001). This might be the practice for various reasons (e.g., data unavailability, computational burden, etc.). Yet, it may not be the most desirable approach since behaviour may display year-to-year variations, and thus, accurate prediction of travel demand might require testing the aforementioned assumption by tracing the changes or stabilities in parameters over time, provided that multiple cross-sectional data are available, which is the focal point of this paper.

Furthermore, building policy-sensitive transportation models adopting an activity-based microsimulation (ABM) framework demands explicit modelling of a full set of activities, including shopping, social or recreational activities, etc., where generation models play a critical role within the framework because their outputs influence the remaining decision-making processes (e.g., mode choice). Hence, in this paper in addition to work- and school-related activities we also examine two aggregate non-work/school (NWS) purposes: marketing/shopping and "other".

The two main objectives of this paper are: (1) to investigate the temporal stability/instability of activity/trip generation behaviour and models associated with work, school, shopping and "other" purposes; and (2) to test year-specific versus "joint/pooled" models of activity/trip generation behaviour over time. Given these objectives, this paper presents a longitudinal analysis of activity generation behaviour in the Greater Toronto and Hamilton Area (GTHA) between 1996 and 2016. The analysis uses the data from the five most recent Transportation Tomorrow Surveys. For work-related activities, behaviour of employed individuals from four occupational sectors are studied separately. Similarly, for school-related activities behaviour of students from three different education levels are examined independently. Further subdivision is made by treating first work/school activity of the day and subsequent work/school activities as distinct activity types. Marketing/shopping activities refer to shopping activities conducted outside the home, i.e., they do not encompass online shopping. All types of out-of-home shopping activities are included in this category, since details concerning the type of shopping activity were not collected in the survey.

"Other" activities include almost[1] every activity that is not collected under work, school or marketing/shopping categories due to the aggregate nature of the purpose definitions in Transportation Tomorrow Survey (TTS), and thus, there is no distinction between "ballet lessons", "visiting grandparents", or "going for drinks with friends" in this study. Yet, from a modelling perspective, the behaviour associated with distinct types of activities within a certain category is expected to be different, not only in terms of activity generation frequency but also in terms of timing, the choice of location and mode of the linked trip. The results of this study confirm this expectation/hypothesis for NWS purposes with respect to activity episode frequency.

Although the findings of this preliminary analysis will serve as stepping-stones in an agent-based modelling context, in particular, for reconstructing the activity generation component of TASHA (Miller and Roorda 2003; Miller et al. 2005), they also provide valuable general insights into the activity/trip generation modelling domain, specifically by investigating the standard "parameter stability" assumption in forecasting over a duration that can be considered a common forecasting period. The rest of this paper is organized as follows. The next section presents a concise overview of literature on longitudinal activity/trip generation behaviour studies that make use of multiple repeated cross-sectional surveys. Section 3 refers to the data used. Section 4 summarizes the methods adopted to investigate the longitudinal activity generation behaviour. Results of the analyses are provided in Sect. 5. Finally, conclusions drawn from this study are reported in Sect. 6.

Literature review

A number of studies have investigated longitudinal activity/trip generation behaviour, making use of multiple cross-sectional data, and using a variety of analysis and modelling methods. Selected important examples of these studies are very briefly discussed here.

Several studies use data sets from two cross-sectional surveys only, testing temporal transferability of models in different geographical regions. For instance, an early temporal transferability study used two household surveys conducted 7 years apart (1964 and 1971) in Indianapolis, to examine the stability of disaggregate trip generation models (Kannel and Heathington 1972). An important conclusion of this study is that the linear regression model estimated from 1964 data successfully predicts the behaviour in 1971. Similarly, Yunker (1976) analyzes the transferability of trip generation models in addition to trip distribution models using two surveys conducted in 1963 and 1972 by the Southeastern Wisconsin Regional Planning Commission. The study finds that the models developed using the former year's data set can predict the behaviour in the latter year with adequate accuracy.

In the GTHA, Badoe and Steuart (1997) use models built for 1964 to predict the trip generation behaviour in 1986 using data from two different surveys and conclude that the models developed for work purposes are more promising than those of shopping purposes. Shams et al. (2014) examine the transferability using two data sets that are 12 years apart, from 1998 and 2010, collected in New York metropolitan region. Similar to Badoe and Steuart's (1997) findings, the researchers indicate that the multinomial logit commuting-trip models are more stable than the models for shopping purposes.

[1] Exceptions include facilitating a passenger or taking children to daycare.

Huntsinger (2012) also investigates the temporal stability of trip generation models in North Carolina with data from years 1995 and 2006. Testing logit models and cumulative logistic regression models mainly using three variables (life cycle, area type, and accessibility), she observes that there is strong evidence of temporal stability for these models and underlines that these models perhaps perform better than the more commonly used cross-sectional classification models. On the other hand, Cotrus et al. (2005) explore trip generation models' temporal transferability estimating two different model specifications: multinomial linear regression and Tobit. Two Israeli National Travel Habit Surveys, from 1984 to 1996/97, are used in the study in which the researchers observe statistical difference in the models from two time periods.

In several other studies, more than two cross-sectional survey data sets are utilized to study longitudinal behaviour. Roorda et al. (2008) estimate disaggregate multivariate regression models comparing travel behaviour trends in Montreal and Toronto, based on various demographic and socio-economic attributes of trip makers using three cross-sectional surveys from each region. The analyses show that there is significant change in the magnitude of the influence of different explanatory variables, such as age and gender, in both regions. In their study on evaluating past mobility patterns to assess potential future patterns in Switzerland, Ciari et al. (2013) conduct a cohort analysis using several Swiss travel diary surveys between 1994 and 2010 and find that the travel behaviour associated with out-of-home activities in Switzerland did not change substantially during the analysis period. Yasmin et al. (2017) investigate the changes in the distributions of attributes that generate activities by analyzing a 10-year period using three independent cross-sectional origin–destination household travel survey data for the Greater Montreal Area. The authors observe significant changes in the frequency distribution of various activities, such as work, school, shopping, etc., by mainly focussing on the impacts of different socio-demographic cohorts on activity generation, however, the study does not include an econometric analysis regarding activity generation.

Similar to the approach that we adopt in our study, common features of the following studies are the use of multiple TTS data sets for the analyses and building models from the pooled data. Mwakalonge and Badoe (2012) make use of TTS data for 1986, 1996, 2001 and 2006, to undertake an empirical investigation using six alternative methods for modelling trip generation. The study results indicate that well-specified models estimated using pooled data from multiple cross-sectional surveys provide trip predictions that have less error than the models estimated based on a single cross-sectional survey data. The researchers also highlight that the pooled models' performances are very comparable to those of the year-specific models. Salem and Habib (2015) and Salem (2016) adopt a multiple discrete–continuous extreme value modelling technique to develop separate activity generation models for workers and non-workers using TTS data for 2001, 2006 and 2011, with the assumption that the individuals maximize their total utility while allocating time to different types of activities considering several time budget-related constraints. Joint models estimated in these studies are found to improve the temporal transferability of activity-based travel demand models.

There are various other important studies that make use of repeated cross-sectional surveys to analyze longitudinal behaviour through investigating transferability of the models, but their focus of analysis is not activity generation. For example, Badoe and Miller (1995, 1998) investigate temporal transferability of mode choice models in the Greater Toronto Area using repeated cross-sectional surveys. Similarly, Badoe and Wadhawan (2002) study mode choice behaviour in the same area using data from 1964 and 1986. Fox (2015) undertakes a thorough investigation on the temporal transferability of mode and destination choice models, conducting analyses for two different regions: Toronto and Sydney.

Since a variety of studies show that building joint models using data from more than a single cross-sectional data enables making better predictions for future years, a longitudinal approach is adopted in this study, and both cross-sectional and joint models are developed to analyze the behaviour over time. Although the activity or trip generation studies listed above analyze longitudinal behaviour benefiting from multiple independent cross-sectional surveys, they do not cover the investigation of the impact of biases introduced into the data owing to the survey methods used to collect the data. This study, on the other hand, also tests the effects of potential survey design biases within the TTS data, along with exploring the activity generation behaviour over a 20-year period. The concept of testing survey design biases is introduced as a secondary objective complementing the two aforementioned main objectives and is explained in detail in Sect. 4.5 of the paper.

Data

The data sources of this study are the five most recent Transportation Tomorrow Surveys: TTS 1996, TTS 2001, TTS 2006, TTS 2011, and TTS 2016. This regional survey has been conducted every 5 years since 1986. It covers almost five percent of the population in and around the Greater Toronto and Hamilton Area. It collects household-based information (e.g., dwelling type, number of vehicles, etc.) along with the attributes of the members of a household (e.g., age, sex, employment status, occupation, etc.) and their trip records (e.g., purpose, location, mode, etc.) from a prior 24-h period on a weekday. The survey is conducted in various languages other than English and French (e.g., Cantonese, Greek, Hindi, Italian, Mandarin, Portuguese, Russian, Spanish, etc.) (Ashby 2018), preventing any language barrier, or associated under-reporting [e.g., Elias et al. (2010)], that might be expected due to having a very diverse environment in the region. Detailed information regarding the survey is available at the Data Management Group's website (Data Management Group—Reports 2014), but summary statistics from the five surveys used in the study are provided in Table 1.

Although the survey has been conducted seven times to date, the data from TTS 1986 and TTS 1991 are excluded from this study. TTS 1986 is excluded owing to the lack of several critical household and survey attributes, whereas TTS 1991 is excluded due to its small sample size relative to the other surveys. Rigorous cleansing was undertaken prior to the analysis to remove inconsistent/incomplete data and filter out unusual behaviour (Ozonder 2018). It is important to note that the study is restricted to analysis of GTHA residents in all survey years, i.e., there is no area that is removed or added in a certain year, the areas from different years overlap fully. The values shown in Table 1 are representative of the GTHA population only, not the complete survey population. Nevertheless, it is important to note that the area outside the GTHA that is covered by the survey constitutes a small share of the total population. Further details regarding the area covered by the survey can be found in the data guides prepared for each survey year which explicitly include a section for "Survey Geography" (Data Management Group 1997, 2003, 2008, 2013a; Ashby 2018).

In addition, TTS data have been supplemented with zone-to-zone distances (e.g., distance between residential location and school location) which were taken from an Emme4 network assignment model for the region (Travel Modelling Group 2015).

Table 1 TTS summary statistics

Attribute/survey	TTS 1996	TTS 2001	TTS 2006	TTS 2011	TTS 2016
Age groups					
[11–17] interval	11%	11%	11%	10%	8%
[18–30] interval	22%	19%	16%	13%	14%
[31–67] interval	58%	59%	59%	59%	59%
[68–98] interval	9%	11%	14%	17%	19%
Mean age (standard deviation)	40 (18)	42 (19)	44 (20)	47 (20)	48 (20)
Sex					
Female	52%	51%	52%	52%	52%
Male	48%	49%	48%	48%	48%
Employment status					
Full-time outside the home	46%	47%	43%	39%	43%
Part-time outside the home	10%	10%	9%	9%	8%
Full-time at home	2%	2%	4%	4%	3%
Part-time at home	1%	1%	1%	1%	1%
Unemployed	42%	40%	43%	47%	44%
Occupation					
General	8%	7%	8%	9%	8%
Manufacturing	13%	13%	9%	7%	7%
Professional	24%	27%	20%	18%	28%
Service	14%	12%	20%	19%	14%
Unemployed	42%	40%	43%	47%	44%
Respondent?					
Yes	43%	42%	42%	43%	45%
No	57%	58%	58%	57%	55%
Average trip frequencies					
Primary work (per worker)	0.79	0.79	0.77	0.75	0.76
Secondary work (per worker)	0.13	0.13	0.12	0.11	0.09
Primary school (per student)	0.79	0.81	0.80	0.79	0.75
Secondary school (per student)	0.02	0.02	0.03	0.02	0.01
Shopping (per individual)	0.19	0.20	0.21	0.23	0.21
Other (per individual)	0.35	0.36	0.34	0.34	0.31
Distribution of trips by purpose					
Primary work	19%	19%	18%	17%	19%
Secondary work	3%	3%	3%	3%	2%
Primary school	7%	6%	6%	6%	5%
Secondary school	0.2%	0.2%	0.2%	0.2%	0.1%
Shopping	8%	8%	9%	10%	10%
Other	15%	14%	14%	15%	14%
Daycare, facilitate passenger, return home	48%	49%	50%	50%	50%
Student status					
Full-time student	17%	17%	17%	15%	13%
Part-time student	3%	3%	2%	2%	2%
Not a student	79%	81%	81%	83%	85%
Mean household size	2.74	2.78	2.72	2.63	2.49

Table 1 (continued)

Attribute/survey	TTS 1996	TTS 2001	TTS 2006	TTS 2011	TTS 2016
Mean number of vehicles in household	1.36	1.45	1.42	1.42	1.43
Survey method					
Phone	100%	100%	100%	83%	34%
Online	–	–	–	17%	66%

The data source is not activity-based per se. Nonetheless, trip-end purposes are used as a proxy for the activity engaged in at the destination. Hence, it has been possible to identify activities from the trip records and develop models to analyze activity generation behaviour.

Methods

Model definitions

Activity generation behaviour for six types of activities is investigated in this study: two types of work-related (first work; second/subsequent work), two types of school-related (first school; second/subsequent school), and two types of NWS-related (marketing/shopping; "other") activities. For work/school-related activities, it is assumed that the determining factors for the first and subsequent occurrences for a certain type of activity might be different, and hence, they are analyzed separately.

A systematic set of steps are followed while conducting the analyses. The general process flow for each activity examined can be summarized as follows:

- Select model structure(s) to be tested.
- Define set(s) of choice alternatives (binomial, trinomial, etc.).
- Choose levels/sectors (sub-categories/samples within the population).
- Estimate year-specific models for all 5 years and assess the outcomes (goodness of fit, signs, order of magnitudes, etc.).
- Estimate four types of joint models, investigate the (scaled) estimates and evaluate the results.

In general, random utility theory-based discrete choice models (McFadden 1978, 1981) are used to explain activity generation behaviour. The error terms that represent the unobserved attributes are assumed to be independently and identically distributed with the Type I Extreme Value distribution, and thus, all the discrete choice models estimated for various activities discussed in this paper are of type logit. For work/school-related activities, multinomial logit (MNL) models are developed. For NWS-related activities, in addition to different discrete choice models (binomial, trinomial, ordered logit), count data models are estimated in this study which are discussed in detail in Sect. 4.4.

For work and school purposes, several other model specifications such as hierarchical/nested, ordered logit, etc. were tested other than the flat logit models. Moreover, for NWS purposes, it might be argued that the choice alternatives are not independent (e.g., a single shopping activity, two or more shopping activities, etc.) violating the independence from irrelevant alternatives (IIA) property (Train 2003) in logit models, and hence, hierarchical model specifications have also been tested. However, these did not generate better fits to the data, nor did their parameter values generally indicate different "behavioural" findings than the models reported here. Furthermore, empirical analyses conducted for NWS purposes showed that count models or ordered logit models did not outperform the flat logit models. Hence, for work/school-related activities we focus herein on the MNL results. Nonetheless, this does not necessarily mean a rejection of any of the other models for possible implementation in a given application.

The standard format of additively separable and linear in parameters systematic utility equation used in the discrete choice models of this study is given in Eq. 1.

$$V_{ijt} = \mu_t\left(\alpha_{jt} + \beta_j x_{ijt}\right) \tag{1}$$

where V_{ijt} = systematic utility for individual "i" for alternative "j" in year "t", α_{jt} = alternative specific constant for alternative "j" in year "t", assumed to be identical for all individuals, μ_t = year-specific scale, assumed to be constant for all the alternatives in a specific year and identical for all the individuals, β_j = vector of parameters for alternative "j", assumed to be identical for all individuals (for year-specific models, different sets of β_j's are estimated), x_{ijt} = vector of explanatory variables for individual "i" for alternative "j" in year "t".

Since repeated cross-sectional data have been used in the analysis, not panel data, the individuals from one survey year are generally different than the set of individuals from another year.

Year-specific logit models are estimated using the maximum likelihood method, mainly using RStudio's (2019) "maxLik" (Henningsen and Toomet 2011) and "mlogit" packages (Croissant 2019). All scales in year-specific models are set equal to 1.0. Further technical details regarding the year-specific logit models are not provided here as they are standard logit models [see Train (2003) for a comprehensive overview]. Models are finalized considering the fits, signs of the parameters, behavioural explanations, etc. after exploring all possible regressors (Montgomery and Runger 2003). Different types of joint models used in this study and their construction methods are described below.

The most general joint model considered is labelled "Joint Model" → JO". In "JO" models, both the scales and alternative-specific constants are allowed to vary from year to year. Thus, keeping 1 year, year 1996, as a base, 4 year-specific scales are estimated in each "JO" model, along with estimating different sets of alternative-specific constants for each year. This application is similar to the approaches adopted in transferability studies. The assumption is that, with the exception of utility function constants and scale, the utility parameters may be temporally invariant. Further details regarding the scale parameter can be found in Badoe (1994), Badoe and Miller (1995, 1998). Three other joint models are also estimated, each of which is a special case relative to the "JO" model:

- *Naïve Pooling → NP:* In this model, it is assumed that both the observed and unobserved factors impacting the behaviour do not change over the years, i.e., the model parameters and the error distribution associated with activity generation remained constant over time [similar to Mwakalonge and Badoe's (2012) Model 3]. A single set of alternative specific constants are estimated for all the years. "NP" models are estimated by pooling the data sets from all 5 years together. Year-specific scales are assumed to be equal to 1, i.e., $\mu_t = 1.0$ where t = {1996, 2001, 2006, 2011, 2016}.

- *Naïve Pooling with Distinct Alternative-Specific Constants→NPDA:* "NPDA" models are very similar to "NP" models, but they assume that the alternative-specific constants vary from year to year. The year-specific scales are assumed to be equal to 1.0.
- *Joint Model with a Single Set of Alternative-Specific Constants→JOSI:* In "JOSI" models, year-specific scales are introduced to account for the variations in parameter estimates over the years. A single set of alternative-specific constants is estimated, similar to the "NP" models.

Note that, excluding the constants and scales, a single vector of parameters is estimated for each alternative in all four types of joint models. Joint model estimations are carried out using the maximum likelihood method, as for the year-specific models. However, the difference in joint models is that the objective function to be maximized is obtained by summing up the log-likelihood functions of each year (Badoe and Miller 1995, 1998). More detailed discussions of the models of the six activity types considered in this paper are presented in the following sub-sections.

Work-related activities

First work-related activity of the day

The population for this activity type is all employed individuals who work full-/part-time either at home or outside the home. This population is further divided into the four occupational categories defined in TTS (Ashby 2018): general office/clerical (sector "G"), manufacturing/construction/trades (sector "M"), professional/management/technical (sector "P"), retail sales and service (sector "S"). All analyses are conducted for each of these categories separately.

Although alternative choice set structures have been tested, such as binomial models (had a work-related activity or not), a trinomial logit model structure is adopted for the analysis of first work-related activity of the day. Separate models for full-time and part-time workers have also been estimated, however, more parsimonious joint models where both worker types are accommodated within a single model are preferred and presented in this paper.

The choices are designated as "A", "B" and "C". "A" represents the choice where the individual does not engage in any work-related activity during the day, either at home or outside the home. Choice "B" means that the individual engages in work-related activity only at home that day and does not go outside the home for any work-related activity. The third alternative, "C", represents the choice where the individual engages in work-related activity outside the home that day, either at their usual workplace or at an alternative location (e.g., a business meeting in a client's office).

There is a question in TTS that is only asked of individuals who work at a workplace other than home, which asks whether the individual conducted any work-related activity at home on the day that the survey is conducted for the case in which they have not made any trips to their usual work-places or for a work-related activity. The individuals who work at home are not eligible for this question. Therefore, in this study it is assumed that if the individuals who work at home have not gone out for a work-related activity on the survey day, then they must have worked at home, and thus, alternative "A" is not made available to the individuals who work at home, either full-time or part-time, because, otherwise it would not be possible to identify the choice that they have made for model estimation purposes since there is no way of knowing how the at-home workers spent their days at home, i.e., whether they spent their time working or not. All three alternatives are made available to the individuals who have a workplace outside the home, as they might have answered the aforementioned survey question as "No" which leads to option "A", if they answer it as "Yes" then it leads to option "B", but if they have gone to their usual workplace or have gone out for a work-related activity then their choice becomes option "C".

Second/subsequent work-related activity of the day

Binomial logit choice models are used to determine if workers who make a first work-related trip in the day make additional work-related trips. The choices are 0 or 1 + subsequent work-related activities, i.e., either the individual does not engage in any further work-related activities in the day or the individual participates in further work-related activities. The exact number of activities in which the individuals participate is not modelled due to the very low frequency of multiple secondary or subsequent work-related activity participation in the survey data from all 5 years. The employed population is used for this activity type. Categorization is made considering different occupational sectors.

School-related activities

First school-related activity of the day

Full- and part-time students are eligible for this activity type. Students are divided into three categories: elementary school students, secondary school students and post-secondary (university/college) students (both undergraduate and graduate). Binary choice models are estimated where the alternatives are going to school and not going to school on the survey day.

Second/subsequent school-related activity of the day

Similar to the second/subsequent work models, binomial logit models are used to predict whether students making a first school trip will make subsequent school-related trips on the survey day. Again, two choices are available: 0 or 1 + trips. And again, the exact number of subsequent activities is not modelled due to the low frequency observed for such cases.

Non-work/school-related activities

The method followed for both NWS activities is the same, hence, is explained under a single heading. The dependent variables in the analyses of marketing/shopping and "other" are the activity episode frequencies. The frequencies represent the number of occurrences of a given type of activity conducted by an individual within a defined period, a day in this case, and thus, count data models[2] are estimated to examine the NWS-related activity

[2] Since the activity participation frequencies are non-negative integers, ordinary least squares regression is not a suitable model structure as it does not restrict the dependent variable to be non-negative (Susmel 2015).

generation behaviour apart from the random utility theory-based discrete choice models as mentioned briefly in Sect. 4.1.

For both marketing/shopping and "other" activities binomial and trinomial logit models are estimated, in addition to ordered logit models. In binomial models the choices available are 0 and 1 +, i.e., no marketing/shopping ("other") activity and one or more marketing/shopping ("other") activities within the survey day. In trinomial models, on the other hand, the choices available are 0, 1, and 2 +, i.e., no marketing/shopping ("other") activity, a single activity of given type, two or more activities of given type within the survey day. Since there has to be at least three ordered choices available to estimate ordered logit models, a trinomial choice set is used in this type of model. Logit models in this modelling exercise do not focus on estimating the exact number of activities participated in due to the very low frequency of multiple NWS activity participation in the survey data from all 5 years.

The frequency distribution of activities has not only influenced the choice structure in logit models but has also affected the choice of count data model types. Count data can be modelled using classical Poisson, geometric and negative binomial regression models (Zeileis et al. 2008), however, the dominance of zeroes in the data sets from 5 years necessitates the use of either zero-inflated models or hurdle models, which are "two-part" models (Susmel 2015). In this study, hurdle models are preferred because zero-inflated models require the outcome to be zero under certain circumstances (e.g., the number of defective items produced in a controlled process is expected to be zero considering a material production problem) (Greene 2011), however, everyone in the surveyed population is eligible for participating in both marketing/shopping and "other" activities. In the *first part* of hurdle type of "two-part" models, the distinction between zero and non-zero is made for the dependent variable ("participation part"), whereas in the *second part*, the non-zero value of the dependent variable is estimated ("amount part"). The two parts of these models are not constrained to contain the same explanatory variables (Susmel 2015). In this study, two types of distributions are tested for the "amount part" of the hurdle models: Poisson and negative binomial. The Poisson distribution is not able to address the problem of over- or under-fitting if there is overdispersion (Ford 2016), i.e., if the variability in the data is greater than what Poisson model can predict, and thus, negative binomial distribution is also tested. The "participation part", on the other hand, is a binomial logit model, and irrespective of the distribution assumed for the "amount part", Poisson or negative binomial, leads to the same likelihood function and hence produces the same parameter estimates in both cases.

Year-specific models are estimated with the aforementioned five model structures: binomial, trinomial, and ordered logits; and hurdle models with Poisson and negative binomial in the *second part*. Since the purpose of estimating count models in this study is to have distinct reference points to compare to logit model structures, not to explore various count data models, their formulations are not included in this paper [see Mullahy (1986), Zeileis et al. (2008), Cameron (2009), Greene (2011), Cameron and Trivedi (2013), Kleiber and Zeileis (2016) for further details]. RStudio's (2019) "pscl" package (Jackman 2017) is used for hurdle model estimations.

Joint models, within which the data are pooled across the years, are estimated using the binomial and trinomial logit structures only. Details regarding the joint models are provided in Sect. 4.1.

In contrast to the analyses conducted for investigating work/school-related activity generation behaviour, the population is not sub-categorized in this preliminary investigation for two reasons. First, the aim is to explore the overall behaviour initially, before segmenting the population. These models lay the foundation for the more detailed/structured models that are required to be developed for operational purposes. Second, the experimental aim of the study on non-work/school purposes was to compare and contrast different modelling methods, to be able to determine the research direction for future studies and to see what would be the differences between adopting the tested approaches (logit models with different choice sets, different count models, etc.).

Testing survey design biases

Controlling for survey bias is a prominent aspect of developing robust travel demand models. Various types of biases in surveys are acknowledged by transportation researchers worldwide (Inbakaran and Kroen 2011; Russell et al. 2004; Richardson 1985). The sources of these biases might be the design of the survey, method of conduct, sample frame, timing of sample selection, non-response, incorrect information, under reporting, etc. (Data Management Group 2013b). Accounting for survey biases is important to prevent inaccuracy in computations, because the models developed will ultimately influence decision-making in traffic, environment and land use-related issues. While some studies focus on proposing solutions/developing methods to overcome or reduce/minimize different types of the aforementioned biases (Chung et al. 2017; Richardson 2003; Stopher and Jones 2001), some studies focus on quantifying/testing the existence of biases. In the Greater Toronto-Hamilton Area (GTHA), Hassounah et al. (1993) investigate the respondent bias using TTS 1986 data and document that there is under-reporting in short non-work/school trips and trips made during off-peak periods. They indicate that the effect of being the respondent was insignificant for home-based work and school trips. Further, Badoe and Steuart (2002) use TTS 1996 data and adopt a multivariate statistical approach to explore the respondent bias in the GTHA, and report that respondent bias is not significant for home-based work/school trips, although they find under-reporting in non-work/school trips, similar to Hassounah et al. (1993).

Furthermore, as noted by Rose (2018), potential survey biases associated with the TTS 2016 survey have not yet been studied. Thus, this study also investigates the impacts of possible survey design biases within five cross-sectional TTS data sets, since the biases are likely to impact the activity frequency estimates, which, in turn, would affect all the remaining model components in an ABM framework. Biases associated with two survey design elements are tested in this study: (1) whether the individual is the respondent[3] within the household or not; and (2) whether the survey was conducted online or on the phone, where the online option has only been available for the two most recent surveys: 2011 and 2016. Hence, this study identifies any biases introduced into the models by these variables.

[3] There can only be one individual from each household who is designated as the respondent, as the survey is completed through a single person, unless the survey is completed online in year 2016.

Results

Work-related activities

First work-related activity of the day

As previously discussed, separate models are developed for the four worker occupation groups defined in the TTS data. Table 2 shows the parameter estimation results[4] for the single-year cross-sectional models for the four occupational sectors.

The variables used in these models are mostly dummy/categorical variables representing different attributes of the individuals. Count of children (age < 11) within the household is the only continuous variable and it represents an attribute of the household. Apart from individual and household attributes, some other factors, such as day of the week, are also tested.

For all the sectors except retails sales and service (sector "S"), it can be seen that individuals are less likely to leave home and conduct a work-related activity outside on Fridays, which probably reflects the inclination towards an extended weekend. It makes sense that this is not the case for the employees in sector "S", as their schedules might be more complicated than the individuals' in other sectors, they might also be working on the weekends, and hence Fridays might not be significantly different than other weekdays. Count of children in the household, on the other hand, does not seem to have an impact on choice probabilities when the employees of sector "M" are considered. This might be an indicator of their less flexible work environment, if they are working at a factory, for instance, they may not be able to stay at home and work. In the other sectors, as the number of children at home increases, the utility of going out to work reduces. With more children at home, the probability of needing to care for dependents for a parent at home might increase, and the obtained results might be reflecting this case. The models show that in general males are more likely to go out to conduct a work-related activity or go outside to work compared to female employees. If an individual has a usual workplace outside the home, s/he is more likely to leave home for a work-related activity on a random weekday. Furthermore, if the individuals who are part-time or full-time students are also part-time employees, their probability of going out to participate in work-related activity reduces. As expected, being a full-time student reduces the utility of alternative "C" more than being a part-time student.

Survey design elements (respondent/non-respondent, survey method (phone, online)) are not found to have significant impacts on choice probabilities for first work-related activity of the day, which is one of the important findings of this study. This implies that there is no considerable under-/over-reporting for this activity even for the non-respondents and for the individuals who filled the survey online.

In terms of goodness-of-fit, ρ^2 values range from 0.18 to 0.40 against constants-only models and 0.51 to 0.66 against null models. When different sectors are compared, it is seen that "Manufacturing" sector has the lowest fit due to its highly heterogeneous nature. Evaluating the trends in ρ^2 values[5] shown in Table 2, it is seen that model fits generally

[4] All parameter estimates are statistically significant at least at a 95% confidence interval and are of expected signs. Parameter names are self-explanatory.

[5] Owing to large sample sizes, adjusted ρ^2 values in all models are almost equal to ρ^2 values, hence they are not provided here.

Table 2 Year-specific models parameter estimates by occupation—first work

Parameters/estimates	Sector "G"					Sector "M"					Sector "P"					Sector "S"				
	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016
B:intercept	−3.37	−2.93	−6.39	−5.17	−4.37	−2.50	−2.19	−4.83	−4.67	−4.48	−2.61	−2.18	−5.74	−5.44	−4.51	−3.13	−3.11	−5.43	−5.08	−4.38
B:sex_male	–	–	–	–	–	−0.65	−0.85	−1.90	−1.43	−0.72	–	–	–	–	–	−0.44	−0.46	−0.93	−0.68	−0.60
B:work_at_home	5.52	4.95	8.61	7.73	7.16	4.01	4.02	7.28	6.90	6.37	3.83	3.61	7.33	7.28	6.76	4.77	4.71	7.34	7.12	6.61
C:count_children	−0.12	−0.14	−0.11	−0.06	−0.09	–	–	–	–	–	−0.09	−0.09	−0.08	−0.04	−0.05	−0.11	−0.07	−0.08	−0.05	−0.06
C:friday.full-/part-time_outside	−0.35	−0.27	−0.39	−0.45	−0.42	−0.27	−0.32	−0.37	−0.39	−0.38	−0.33	−0.29	−0.30	−0.39	−0.40	–	–	–	–	–
C:full-time_student.work part-time	−1.39	−1.12	−1.11	−1.43	−1.36	−1.23	−1.27	−1.23	−1.70	−1.47	−1.11	−1.03	−0.89	−1.19	−0.95	−1.37	−1.39	−1.31	−1.35	−1.49
C:intercept	0.84	0.73	0.75	0.71	0.75	0.91	1.03	0.91	0.92	0.81	0.63	0.64	0.59	0.57	0.61	0.70	0.69	0.69	0.54	0.59
C:part-time_student.work part-time	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	−0.49	−0.56	−0.28	−0.43	−0.52
C:sex_male	0.12	0.10	0.13	0.10	0.07	–	–	–	–	–	0.17	0.25	0.18	0.16	0.17	0.25	0.23	0.04	0.10	0.13
C:work_full-time_outside	1.40	1.52	1.53	1.43	1.57	1.24	1.17	1.24	1.27	1.43	1.41	1.41	1.41	1.37	1.59	1.10	1.11	1.13	1.13	1.13

Table 2 (continued)

Parameters/estimates	Sector "G"					Sector "M"					Sector "P"					Sector "S"				
	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016
ρ^2 against constants-only model	0.21	0.24	0.33	0.37	0.40	0.18	0.18	0.24	0.25	0.28	0.20	0.22	0.33	0.37	0.39	0.25	0.26	0.34	0.36	0.38
ρ^2 against null model	0.61	0.59	0.65	0.63	0.65	0.61	0.61	0.64	0.64	0.65	0.59	0.58	0.62	0.61	0.66	0.52	0.51	0.55	0.54	0.54
Number of observations	16,005	19,113	21,208	22,978	20,038	23,781	30,642	20,533	15,841	13,572	47,009	70,047	51,030	46,004	67,567	26,486	30,919	49,484	48,672	32,304

Not working is denoted by choice "A" (base), working at home is represented by choice "B", choice "C" indicates working outside improve over time, which begs the questions: (1) Is the survey picking more homogeneous people over time (is there a selectivity bias?)? (2) Are people getting more and more systematic over time, displaying less variability due to increased stress, congestion, etc.? (3) Or is it both?

The joint models for this activity are developed by using the same variables from the year-specific models for each occupational category with the adjustments described in Sect. 4.1, Model Definitions. The outputs of the models are shown in Table 3. For each sector, parameter estimates are generally close to the values obtained from year-specific models. This holds for all four joint model types since different types of joint models have not led to significantly different parameter estimates.

All the year-specific scales are close to 1.0 with the highest value being less than 1.5 and the lowest value being more than 0.95. In general, there is a slightly upward trend in scales over the years in all sectors, indicating that the variance is getting lower. This, again, suggests that individuals' behaviour is getting more systematic, less random over time, and/or perhaps the surveys are picking more similar individuals.

Almost the same goodness-of-fit values are obtained from the four different joint models in each sector; i.e., they appear to perform equally well. Models of sectors "G" and "S" perform better ($\rho^2 > 0.32$) than the models of sector "P" ($\rho^2 \sim 0.30$), and the models of sector "M" have the lowest fit ($\rho^2 \sim 0.21$).

When the year-specific alternative specific constants from the "NPDA" models are averaged, the values obtained are very close to the parameter estimates for the intercepts from the "NP" models. This holds for both alternatives "B" and "C", where alternative "A" is the base. Similarly, when the intercepts from "JOSI" and "JO" models are scaled and then averaged, close results are obtained. This suggests that the joint models are rather similar. However, there is no clear trend in year-specific intercepts of the models. The portion of information that is not captured by the models varies from year to year, which might be attributed to the survey characteristics and having independent surveys.

When the parameter estimates from "JO" models are scaled, the variation of estimates shows similar patterns that are observed when the year-specific models' estimates are plotted. In the case of scaled estimates from "JOSI" models, smoother patterns are observed, since in "JOSI" models unobserved attributes from different years are accumulated in a single intercept value for each alternative. Figures showing the plots of scaled estimates can be found in the supplementary material (Figures 1.1.2 and 1.1.3).

The year-specific scales estimated for "JOSI" models are usually lower than the values obtained from "JO" models, which explains the differences in the order of magnitude of coefficient of variations calculated to analyze the variation in estimates from year to year, because it is mathematically shown that the coefficient of variation of parameter estimates in "JOSI" and "JO" models does not depend on the estimated value (except for the intercepts in "JO" models), rather, it depends on the scales and the total number of time periods (see the supplementary material: Part 1.3). The lower the coefficient of variation the lower the dispersion of the values around the mean (Insee 2016). In this study it is found that, for all sectors, the intercepts in the "JO" models have higher coefficients of variation when compared to the other parameter estimates for first work-related activity of the day. Nevertheless, computed coefficient of variation values for other parameter estimates are less than 20%, and for the intercepts of "JO" models the highest value is less than 40%. To further investigate the temporal stability of parameters, when the estimates from different years are plotted on a single graph for each occupational category separately (refer to the supplementary material—Figure 1.1.1), two main conclusions can be drawn for all four sectors. First, it can be seen that the variation in the estimated values over time are negligible. The

Table 3 Joint models parameter estimates by occupation—first work

Parameter estimates	Sector "G"				Sector "M"				Sector "P"				Sector "S"			
	NP	NPDA	JOSI	JO	NP	NPDA	JOSI	JO	NP	NPDA	JOSI	JO	NP	NPDA	JOSI	JO
B:intercept	−3.85	–	−3.80	–	−2.74	–	−2.68	–	−3.12	–	−3.01	–	−3.91	–	−3.80	–
B:intercept_1996	–	−3.73	–	−3.54	–	−2.70	–	−2.47	–	−3.24	–	−2.77	–	−3.71	–	−3.39
B:intercept_2001	–	−3.45	–	−3.31	–	−2.55	–	−2.32	–	−2.90	–	−2.44	–	−3.75	–	−3.44
B:intercept_2006	–	−4.37	–	−4.18	–	−3.20	–	−3.12	–	−3.50	–	−3.29	–	−4.24	–	−3.89
B:intercept_2011	–	−4.12	–	−3.95	–	−3.01	–	−3.01	–	−3.32	–	−3.17	–	−3.99	–	−3.70
B:intercept_2016	–	−3.78	–	−3.64	–	−2.69	–	−2.72	–	−2.89	–	−2.85	–	−3.77	–	−3.50
C:intercept	0.75	–	0.73	–	0.93	–	0.90	–	0.61	–	0.56	–	0.63	–	0.60	–
C:intercept_1996	–	0.77	–	0.84	–	0.93	–	1.05	–	0.59	–	0.82	–	0.73	–	0.76
C:intercept_2001	–	0.79	–	0.93	–	0.97	–	1.13	–	0.66	–	0.97	–	0.73	–	0.75
C:intercept_2006	–	0.78	–	0.63	–	0.89	–	0.61	–	0.57	–	0.34	–	0.66	–	0.57
C:intercept_2011	–	0.64	–	0.51	–	0.91	–	0.56	–	0.49	–	0.27	–	0.53	–	0.44
C:intercept_2016	–	0.78	–	0.62	–	0.94	–	0.58	–	0.71	–	0.37	–	0.57	–	0.47
B:sex_male	–	–	–	–	−1.00	−0.95	−0.97	−0.87	–	–	–	–	−0.66	−0.65	−0.64	−0.58
B:work_at_home	6.30	6.44	6.18	5.98	4.59	4.65	4.47	4.35	4.84	4.88	4.63	4.32	5.86	5.91	5.68	5.38
C:count_children	−0.11	−0.11	−0.10	−0.10	–	–	–	–	−0.07	−0.07	−0.06	−0.06	−0.07	−0.07	−0.06	−0.06
C:friday.full-/part-time_outside	−0.38	−0.38	−0.37	−0.35	−0.33	−0.33	−0.32	−0.29	−0.34	−0.34	−0.32	−0.27	–	–	–	–
C:full-time_student.work part-time	−1.25	−1.26	−1.22	−1.19	−1.30	−1.30	−1.26	−1.21	−1.01	−1.02	−0.95	−0.84	−1.35	−1.38	−1.30	−1.24
C:part-time_student.work part-time	–	–	–	–	–	–	–	–	–	–	–	–	−0.43	−0.44	−0.41	−0.39
C:sex_male	0.10	0.10	0.09	0.09	–	–	–	–	0.19	0.19	0.18	0.15	0.13	0.13	0.12	0.11
C:work_full-time_outside	1.49	1.49	1.45	1.36	1.24	1.25	1.20	1.07	1.45	1.44	1.36	1.15	1.13	1.12	1.08	1.00
scale_2001	–	–	0.97	0.95	–	–	1.01	0.98	–	–	0.98	0.95	–	–	0.99	1.00
scale_2006	–	–	1.06	1.16	–	–	1.06	1.30	–	–	1.09	1.38	–	–	1.07	1.15
scale_2011	–	–	1.01	1.16	–	–	1.07	1.36	–	–	1.07	1.40	–	–	1.04	1.18
scale_2016	–	–	1.08	1.17	–	–	1.09	1.37	–	–	1.18	1.46	–	–	1.07	1.19
ρ^2 against constants-only model	0.323	0.324	0.319	0.321	0.210	0.211	0.211	0.213	0.302	0.303	0.301	0.305	*0.327*	*0.329*	*0.325*	*0.327*

Table 3 (continued)

Parameter estimates	Sector "G"				Sector "M"				Sector "P"				Sector "S"			
	NP	NPDA	JOSI	JO	NP	NPDA	JOSI	JO	NP	NPDA	JOSI	JO	NP	NPDA	JOSI	JO
ρ^2 against null model	0.627	0.628	0.628	0.629	0.626	0.627	0.627	0.628	0.607	0.608	0.608	0.610	*0.531*	*0.532*	*0.531*	*0.532*
Number of observations	99,342				104,369				281,657				187,865			

values remain rather stable over a 20-year period for most of the parameters, such as the dummy for conducting the activity on Friday, the dummy for sex (female being the base category), or the dummy for student status of part-time workers. Second, the parameter for the dummy representing the working at home employment status for individuals seems to display slightly more variation over time. This indicates some statistical fluctuation in propensity for working at home. However, the summation of the intercept and work at home dummy seems to be relatively stable over time, which might be an indicator that if the individual is employed at home the net impact on the utility of alternative B does not change much over the years.

Both of these findings are in fact parallel to the findings stated in Miller et al. (1998) regarding the Greater Toronto Area trip generation rates between 1986 and 1996. The researchers investigate the trip rates over 10 years in almost the same region and conclude that the work trip rates per worker employed outside the home remain fairly constant over time. In addition, they foresee that possible future changes might be attributable to the changes in the labour participation rate of working at home, rather than trip rates per se.

Aforementioned findings are further supported by statistical tests conducted. Table 4 shows t values obtained from the tests conducted to examine the temporal statistical equivalence of the year-specific model parameters and the corresponding "NP" type of joint model parameters. The idea is to test whether the full time-series of data can be exploited without loss of information to build robust models, since statistical equivalence of year-specific model parameters and joint model parameters would support this argument. The "NP" type of joint model is chosen for comparison due to its consistency with the year-specific model structure which does not include any year-specific intercepts or scales.

The null hypothesis of statistical equivalence, i.e., the difference between the parameters being equal to zero, cannot be rejected at a 95% confidence interval for most of the parameters in most of the years since the t value is between − 1.96 and + 1.96 (values in "bold" format), indicating that there is considerable stability in model parameters over time. The null hypothesis is rejected mostly for the intercept of alternative B, i.e., staying at home to work, and the dummy of "work at home" employment status.

No systematic patterns were observed in the pairwise t-tests conducted to test the equivalence of the year-specific model estimates over time considering all yearly combinations either, apart from the previously mentioned ones. The t values obtained from these tests can be found in the supplementary material (Tables 1.2.1–1.2.4), where, again, the values in "bold" show that the null hypothesis of equality cannot be rejected at a 95% confidence interval. In the tables, column headings show the 2 years used in a certain combination.

All in all, although there is slight fluctuation in estimates, no strictly increasing/decreasing trend is observed over the years across the sectors. The range of the values taken by the estimates are very narrow in most cases, and thus, their contribution to the utility remains stable. The minor fluctuations observed may be a side-product of having independent survey samples each year.

Second/subsequent work-related activity of the day

Participation in second/subsequent work-related activities are similarly modelled by occupation group. For the interested reader, model outcomes are discussed in the supplementary material and the parameter estimates for the year-specific models and joint models are shown in Tables 2.1 and 2.2 respectively. The models do not have strong explanatory powers and they are not discussed here in detail. Nevertheless, for the second/subsequent

Table 4 Temporal statistical equivalence test results—t values—first work

t values	Sector "G"					Sector "M"					Sector "P"					Sector "S"				
	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016	1996	2001	2006	2011	2016
B:intercept	3.92	9.36	−3.40	−5.19	−2.77	2.02	5.52	**−1.50**	−4.05	−5.18	9.47	22.50	−9.93	−10.77	−11.56	9.10	10.25	−9.54	−7.84	−3.78
B:sex_male	–	–	–	–	–	2.82	**1.40**	−3.28	**−1.79**	**1.46**	–	–	–	–	–	2.16	2.15	−3.46	**−0.17**	**0.75**
B:work_at_home	−4.07	−9.00	3.07	5.24	4.04	−4.49	−4.81	**1.73**	4.54	5.44	−12.90	−19.85	9.30	11.08	14.68	−10.04	−11.59	9.04	8.22	5.47
C:count_children	**−0.54**	**−1.24**	**−0.32**	**1.45**	**0.37**	–	–	–	–	–	**−1.19**	**−1.57**	**−0.37**	2.02	**1.39**	−2.23	**0.03**	**−0.56**	**1.03**	**0.19**
C:friday.full-/part-time_outside	**0.50**	2.18	**−0.20**	**−1.54**	**−0.84**	**1.33**	**0.38**	**−0.71**	**−0.99**	**−0.78**	**0.20**	**1.66**	**1.17**	**−1.59**	−2.10	–	–	–	–	–
C:full-time_student.work part-time	**−1.50**	**1.42**	**1.37**	**−1.57**	**−1.00**	**0.62**	**0.29**	**0.42**	**−1.94**	**−0.99**	**−0.86**	**−0.12**	**1.25**	**−1.44**	**0.67**	**−0.43**	**−0.82**	**0.99**	**−0.02**	−2.88
C:intercept	**1.60**	**−0.39**	**−0.11**	**−0.91**	**0.05**	**−0.43**	**1.57**	**−0.28**	**−0.20**	**−1.51**	**0.55**	**0.86**	**−0.36**	**−0.94**	**0.14**	**1.89**	**1.79**	**1.94**	−3.17	**−1.13**
C:part-time_student.work part-time	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	**−0.59**	**−1.42**	**1.79**	**−0.04**	**−0.93**
C:sex_male	**0.35**	**0.08**	**0.56**	**−0.01**	**−0.57**	–	–	–	–	–	**−0.60**	2.50	**−0.24**	**−1.26**	**−0.86**	3.28	3.03	−3.15	**−1.09**	**0.19**
C:work_full-time_outside	**−1.61**	**0.61**	**0.72**	**−1.12**	**1.35**	**−0.14**	**−1.26**	**−0.10**	**0.34**	2.14	**−0.97**	**−0.98**	**−1.01**	**−1.90**	3.57	**−0.57**	**−0.47**	**0.12**	**0.14**	**0.22**

work-related activities, statistically significant respondent bias has been found in the models which is a noteworthy result, implying that the respondents may not be aware of the activities taking place in another household member's schedule later in a day when compared to being more informed about the first work-related activity. Lastly, the variations in parameters estimates are more than what is observed for first work-related activity, but still, fluctuations generally remain in a narrow band.

School-related activities

First school-related activity of the day

First school-related activity generation (e.g., attending class, participating in a workshop, attending a seminar, etc.) is modelled separately for elementary, secondary and post-secondary (university/college) students. The only explanatory variable that is found to be statistically significant in the elementary and secondary school models is the "Friday" dummy, which reduced the utility of going to school on Fridays. I.e., the alternative specific constants provide virtually all explanatory power in these models—essentially a simple fixed average trip rate model. For the sake of conciseness, the results for those models are not provided in further detail in this paper. Model results for post-secondary students are shown in Table 5 (year-specific models) and Table 6 (joint models).

It can be seen that, again, the parameter estimates remain relatively stable over time for all variables. There are some fluctuations, but their order of magnitude is not large. The students from the 18 to 30 age group are more likely to go to school on a random weekday when compared to older generations, however, if the 18–30 group is disaggregated further it can also be seen that the increase in utility reduces as the age increases. The number of babies in the household, the distance between the residential zone and school zone, and the day of the week being Friday reduce the probability of going to school. Being a full-time student considerably increases the likelihood of going to school compared to being a part-time student. Having a full-time job outside the home decreases the utility of taking part in a school-related activity. Different occupations have varying impacts on the probabilities. Being employed in sectors "G" and "P" reduces the utility of going to school more than being employed in sectors "M" and "S". Males are more likely to go to school on a random weekday than females. It is crucial to note that the survey design elements have not been found to have significant impacts on choice probabilities for the first school-related activity of the day.

Unlike the trend observed in the first work-related activity models, school model performance improves as one goes back in time. This might indicate that the behaviour is getting more random over the years, and perhaps, the models are not able to capture the variations as well. It might also indicate that the samples chosen for the survey may not be as representative of the student population.

The parameter estimates obtained for the four joint models are similar for most of the variables (Table 6). Fluctuations are observed in the estimates of dummies for different occupational sectors, mainly in sector "M". There is a decreasing trend in the year-specific intercepts of the "NPDA" and "JO" types of joint models. The trend is continued to be observed when the estimates of "JO" model is scaled. Since not going to school is the base alternative, this trend suggests that university/college students are becoming less likely over time to go to school on a random weekday, which might be the case for various reasons.

Table 5 Year-specific models parameter estimates—first school

Post-secondary school students					
Parameters/estimates	1996	2001	2006	2011	2016
1:intercept	0.00	0.06	−0.11	−0.40	−0.31
1:age_18-21	0.57	0.47	0.55	0.52	0.16
1:age_22-25	0.17	–	0.24	0.19	–
1:age_26-30	–	–	0.21	0.17	–
1:count_babies	−0.27	−0.30	−0.23	−0.33	−0.29
1:distance.home_school_zone	−0.02	−0.02	−0.01	−0.01	−0.01
1:friday	−0.66	−0.64	−0.63	−0.58	−0.49
1:full-time_student	1.61	1.52	1.41	1.50	1.42
1:full-time_worker.outside	−0.69	−0.70	−0.61	−0.59	−0.86
1:occupation_G	−0.40	−0.43	−0.54	−0.47	−0.43
1:occupation_M	−0.32	−0.30	−0.47	−0.35	−0.19
1:occupation_P	−0.51	−0.57	−0.84	−0.70	−0.63
1:occupation_S	−0.27	−0.21	−0.33	−0.37	−0.19
1:sex_male	0.16	0.10	0.13	0.13	0.12
ρ^2 against constants-only model	0.25	0.22	0.20	0.18	0.16
ρ^2 against null model	0.29	0.27	0.23	0.20	0.17
Number of observations	19,285	21,975	20,416	18,200	16,432

Some advancements in technology in terms of "online courses/teaching" might have been altering school-related travel behaviour. For instance, in various studies the effect of online schooling on carbon emissions have been studied (Caird et al. 2014; Roy et al. 2008; Versteijlen et al. 2017). In some of these studies it has been found that online learning methods can highly reduce emissions, and in some other studies it has been highlighted that there is a strong potential in achieving a more sustainable environment through these innovative methods. Given these (potential) benefits and considering the increasing environmental concerns globally along with the need to take action, these types of courses might be more promoted in the future and we might be able to observe their impacts. Yet, the aforementioned studies do not examine behaviour over a long period. The longitudinal impacts of improvement in technology on student travel behaviour to school remains a topic that might be investigated by researchers in the future. In addition, carbon emissions impacts is only one of the many dimensions which form the debate about the benefits/drawbacks of online teaching.

Considering the case of the GTHA, it can be noted that the universities/colleges in the region provide the opportunity of taking certain classes online which provides the flexibility of following the course without location/timing requirements, i.e., one neither has to be in the classroom nor has to listen to the lecture at a specific time-period, students can watch the lecture videos, complete course-related exercises, assignments, etc. when and where they are available. It might be argued that the online course option might have reduced the rate of school trips. Nevertheless, usually a post-secondary school student does not take all the courses online, thus, needs to attend the classes at school in addition to going to school for other gatherings such as meetings, presentations, seminars, etc. In this study, the impact of online courses cannot be quantified due to data unavailability. Hence, this aspect is not taken into account while comparing the stability of models over time. Nonetheless, it

Table 6 Joint models parameter estimates—first school

Post-secondary school students				
Parameter estimates	NP	NPDA	JOSI	JO
1:intercept	−0.09	–	−0.10	–
1:intercept_1996	–	0.10	–	0.07
1:intercept_2001	–	0.05	–	0.05
1:intercept_2006	–	−0.12	–	−0.13
1:intercept_2011	–	−0.27	–	−0.29
1:intercept_2016	–	−0.38	–	−0.37
1:age_18-21	0.39	0.38	0.45	0.41
1:count_babies	−0.28	−0.30	−0.32	−0.31
1:distance.home_school_zone	−0.01	−0.01	−0.02	−0.01
1:Friday	−0.60	−0.60	−0.68	−0.64
1:full-time_student	1.52	1.53	1.69	1.62
1:full-time_worker.outside	−0.69	−0.70	−0.75	−0.73
1:occupation_G	−0.42	−0.44	−0.30	−0.27
1:occupation_M	−0.27	−0.33	−0.70	−0.69
1:occupation_P	−0.64	−0.65	−0.30	−0.34
1:occupation_S	−0.27	−0.26	−0.48	−0.47
1:sex_male	0.13	0.13	0.14	0.14
scale_2001	–	–	0.95	0.95
scale_2006	–	–	0.89	0.94
scale_2011	–	–	0.84	0.95
scale_2016	–	–	0.74	0.86
ρ^2 against constants-only model	0.200	0.204	0.199	0.202
ρ^2 against null model	0.230	0.234	0.232	0.234
Number of observations	96,308			

can be noted that this might be one of the factors that we are not able to capture over time which leads to the decreasing model fit trend we observed in the post-secondary school models.

The joint model results show that all year-specific scales are less than 1.0, with an overall decreasing trend in time, which might suggest that the variation is getting higher in time. The goodness of fit obtained from the joint models are almost the same for all models, with ρ^2 values of approximately 0.20 against a constants-only model and 0.23 against a null model. The reason why the ρ^2 values calculated using the likelihood of a constants-only model and the likelihood of the null model are very close is that the ratio of 0 s and 1 s are close to 1 in each year, which is equal to the probability of each alternative when there is no model.

Table 7 shows the t values obtained from the tests conducted to investigate the temporal statistical equivalence of the year-specific model parameters and the corresponding "NP" type of joint model parameters for first school-related activity of the day for post-secondary school students. The null hypothesis of statistical equivalence cannot be rejected at a 95% confidence interval for most of the parameters in most of the years since the t value is between −1.96 and + 1.96 (values in "bold" format), showing that there is considerable stability in parameters. No systematic patterns were observed when the rejections

Table 7 Temporal statistical equivalence test results—t values—first school

Post-secondary school students					
t values	1996	2001	2006	2011	2016
1:intercept	**1.73**	2.85	**−0.31**	−5.62	−3.91
1:age_18-21	3.50	**1.90**	2.99	2.41	−5.45
1:count_babies	**0.17**	**−0.20**	**0.57**	**−0.43**	**−0.01**
1:distance.home_school_zone	−3.17	**−1.06**	**−0.34**	3.80	4.80
1:friday	**−1.38**	**−1.03**	**−0.61**	**0.48**	2.52
1:full-time_student	**1.76**	**0.12**	−2.09	**−0.21**	**−1.90**
1:full-time_worker.outside	**−0.08**	**−0.20**	**1.12**	**1.17**	−2.21
1:occupation_G	**0.30**	**−0.03**	**−1.42**	**−0.56**	**−0.09**
1:occupation_M	**−0.58**	**−0.44**	**−1.95**	**−0.62**	**0.62**
1:occupation_P	**1.85**	**1.08**	−2.86	**−0.77**	**0.13**
1:occupation_S	**−0.06**	**1.26**	**−1.48**	−2.17	**1.68**
1:sex_male	**0.93**	**−0.68**	**−0.03**	**−0.03**	**−0.08**

are considered. These observations are also valid for the tests where the yearly pairwise combinations of estimates are inspected (Refer to Table 3.1 in the supplementary material).

Second/subsequent school-related activity of the day

This analysis only covers post-secondary students; elementary and secondary school students make very few second/subsequent school trips. Similar to the second/subsequent work-related activity, this activity is conditional on having a first-school activity in the day. The outputs of the cross-sectional models and joint models estimated for the second or subsequent school-related activity of the day are shared in the supplementary material for the interested reader (Tables 4.1 and 4.2. respectively). Despite the poor performance of the models, it should be noted that significant respondent bias has been found for this activity and in general parameter estimates exhibit stability.

Non-work/school-related activities

Marketing/shopping

For marketing/shopping activities five types of cross-sectional models are developed using the same explanatory variables[6]: binomial logit, trinomial logit, ordered logit, hurdle with Poisson, hurdle with negative binomial, for which the results were very similar. Hence, only the results of binomial logit model are shown here (Table 8), the results of the remaining models can be found in the supplementary material (Tables 5.1–5.3). By the same

[6] Note that, insignificant variables were removed from the second part of the count data models after exploring all possible regressors.

Table 8 Year-specific models parameter estimates of binomial logit model—marketing/shopping

Binomial logit parameter estimates	1996	2001	2006	2011	2016
1+:intercept	−1.35	−1.42	−1.42	−1.36	−1.18
1+:count_TRE	−0.09	−0.05	−0.05	−0.06	−0.13
1+:friday	0.21	0.22	0.19	0.19	0.19
1+:full-time_student	−1.22	−1.33	−1.38	−1.30	−1.41
1+:income_above_100K	–	–	–	–	0.13
1+:income_below_15K	–	–	–	–	−0.33
1+:online	–	–	–	–	−0.14
1+:reside_in_PD1	−0.68	−0.74	−0.52	−0.61	−0.23
1+:reside_in_PD2_6	−0.29	−0.25	−0.19	−0.16	−0.07
1+:respondent	0.59	0.66	0.64	0.62	0.64
1+:sex_male	−0.22	−0.21	−0.17	−0.17	−0.12
1+:suff_veh	0.45	0.45	0.46	0.46	0.39
1+:work_home_full-time	−0.46	−0.46	−0.40	−0.36	−0.47
1+:work_outside_full-time	−0.79	−0.85	−0.85	−0.87	−1.15
ρ^2 against constants-only model	0.07	0.08	0.08	0.08	0.09
ρ^2 against null model	0.42	0.41	0.41	0.37	0.42
Log-likelihood	−81,361	−108,580	−109,150	−117,670	−108,500
AIC	162,743	217,187	218,314	235,357	217,028
Number of observations	203,931	267,683	264,845	271,457	267,798

token, the results from different tables are not discussed separately for brevity, as the same conclusions can be drawn from them.

The variables used in these models are mostly dummy/categorical variables representing different attributes of the individuals or attributes of their households. It can be seen that individuals are more likely to shop on Fridays. In addition, if the individual owns a driver's license and the number of vehicles owned by their household is more than or equal to the total number of driving licenses in the household then s/he has higher probability of shopping, which might be an indicator of ease of access to shopping locations.

The results further indicate that with an increase in the number of trip-record-eligible (TRE) individuals[7] in a household, the probability of the modelled individual's participation in a marketing/shopping activity reduces, which is intuitive. Full-time students and individuals who work full-time, either at home or outside the home, are less likely to be shopping on a random weekday. When the parameter estimates of these variables are compared, it can be concluded that the utility of shopping decreases the most if the individual is a full-time student and decreases the least if the individual works full-time at home, which makes sense, as work-at-home individuals' schedules might be more flexible when compared to the others. Also, if there are non-student individuals in a household, they may be more likely to shop when compared to the full-time students due to availability of time, income (given they are employed), etc. Based on the results, it can be deduced that males have a lower tendency to shop.

[7] Individual's age should be greater than or equal to 11 to be considered as "TRE". This definition follows from the fact that TTS does not record the trips of individuals younger than 11.

Furthermore, in TTS 2016, households' total income class data were collected for the first time (Ashby 2018) and the income classes were used as variables in the 2016 models. The results show that individuals from a higher household income class (above $100,000 annually) are more likely to shop on a random weekday, whereas the individuals from lower income class-households (less than $15,000 annually) have a lower probability of going shopping.

"reside_in_PD1" and "reside_in_PD2_6" variables have negative parameter estimates. However, this does not necessarily imply that one is less likely to shop if s/he resides in PD[8] 1-6. PD 1 covers downtown Toronto, and PD 2-6 are the closest neighbouring planning districts. These urban areas are densely-packed with various establishments, residential units, firms, etc., where residents are more likely to be making shopping trips by walking, and it is very likely that most of these trips have not been fully captured by the survey (Harding et al. 2018).

When the survey design elements are tested, it is seen that there is a significant respondent bias in all years, revealing that there is an important under-reporting issue for the non-respondents. This finding is parallel to previous studies' findings (Hassounah et al. 1993; Badoe and Steuart 2002). Moreover, in 2016, the survey method is also found to be significant, indicating that marketing/shopping trips are under-reported when the survey is completed online. This might be because of survey fatigue, as one has to enter all the trips and their associated details herself/himself in the online version.

It is crucial to note that with the advancements in technology, the nature of shopping has been reshaped. Now, one can shop online without having to go to the store. This has major implications in terms of travel demand/behaviour which have been investigated by various studies in the literature (e.g., Farag et al. 2007; Winslott Hiselius et al. 2015; Lee et al. 2017; Weltevreden and Van Rietbergen 2007). While some studies find that online shopping leads to reduced physical shopping, some other studies find that there is no decrease in physical shopping rates despite the availability of online shopping. For example, in a study conducted in the Netherlands, Weltevreden and Van Rietbergen (2007) find that over 20% of online shoppers made less trips to the city centre due to e-shopping. On the other hand, in Sweden, Winslott Hiselius et al. (2015) find that individuals who prefer frequent online shopping make as many trips to a store as less frequent online shoppers, underlining that the time saved from online shopping is used for additional physical shopping in addition to running other errands. Lee et al. (2017) investigate the relationship between online and traditional in-store shopping frequency of Davis (California) residents and conclude that shopping online is in fact associated with higher rates of in-store shopping.

There is also another dimension to the concept of online shopping which covers the changes in the amount of freight transport (e.g., Golob and Regan 2001; Rotem-Mindali and Weltevreden 2013). However, since we are focussing on individuals' activity/travel behaviour in this paper, this aspect is not discussed.

In this study, unfortunately, we do not have detailed information on the shopping episodes in the survey data. For example, as previously indicated, we do not have any information on the type of shopping (e.g., grocery, clothing, etc.), which makes it even more complicated to investigate the impacts of online shopping and compare the changes in behaviour over time. Furthermore, the fits of the models are lower than reasonable model fit values signalling that there are aspects of this behaviour that cannot be well-captured

[8] "PD" stands for "planning district".

with the estimated models. This, in turn, has implications in terms of data collection. To be able to quantify the impacts of technological improvements on the activity/travel behaviour and to be able to update our operational models to incorporate these changes enabling them to be more policy-responsive (e.g., Suel and Polak 2018), we need to revise our surveys to be more activity-oriented, rather than collecting a summary of trip diaries solely.

Despite the general stability in parameter estimates in the models of marketing/shopping activities, the estimates for some variables (e.g., reside_in_PD1, reside_in_PD2_6, work_outside_full-time, etc.) in the year 2016 seem to show some divergence from the remaining years even though the yearly distributions of the corresponding variables do not reflect such a trend. This might be an indication of the beginning of changes in behaviour. Yet, the factors influencing this change cannot be quantified with the data available. The changes might be attributed to the aforementioned technological developments. For example, the parameter estimate for workers who work outside the home full-time has become more negative in 2016, implying a further reduction in the utility of shopping outside. This might be associated with the fact that these individuals might be more occupied, and hence, can spend less time on outside shopping, when compared to unemployed, part-time workers, etc., and that is why, perhaps, they are switching to e-shopping. This begs an important question: will we be observing more changes in the upcoming survey in 2021, or will the dissimilarities in 2016 remain as a statistical fluctuation? This issue remains to be tested in the future until the results of the next survey become available.

In terms of model performances, it should be underlined that the models are not able to explain the variability in behaviour very well, i.e., low fit values are obtained ($\rho^2 \sim 0.07–0.08$ in logit models). This might be attributed to several reasons: (1) there is more variability in marketing/shopping activity generation behaviour when compared to more standardized commitments such as work/school-related activities; (2) the activity type is heterogeneous by definition, i.e., it is not possible to capture the difference between "buying milk" and "buying a computer", where the underlying behaviour is inherently different; (3) lack of accessibility and preference/taste-related explanatory variables; (4) having only a single day's information from a person's schedule for an activity which potentially has a longer planning range, which might better modelled on a weekly (or longer) basis (Dianat et al. 2018).

When the log-likelihood (LL), Akaike Information Criterion (AIC), etc. of the models estimated for a certain year are compared, it is seen that binomial logit models perform better than trinomial and ordered logit models, which perform equally well/poorly.[9] The count data models perform slightly worse than the logit model specifications with lower log-likelihood values. This ordering of performance between different model structures holds in all 5 years. However, it should be noted that the slightly better performance of the binomial logit models is likely due to the aggregation of available choices into two alternatives compared to three. With fewer alternatives, the variability that needs to be explained by the model reduces, hence, the binomial models provide higher fit values. The same explanation is valid for the lower fits of count data models, as they are designed to predict the exact count of activities participated in, leading to more alternatives, which, in fact, is the advantage of using count data models. Though the performances of count data models are very similar, tests on dispersion show that hurdle models that utilize negative binomial

[9] This is expected as they are almost the same model in the given particular case.

Table 9 Binomial joint models parameter estimates—marketing/shopping

Binomial logit	NP	NPDA	JOSI	JO
1+:intercept	−1.37	–	−1.37	–
1+:intercept_1996	–	−1.39	–	−1.39
1+:intercept_2001	–	−1.34	–	−1.28
1+:intercept_2006	–	−1.35	–	−1.31
1+:intercept_2011	–	−1.31	–	−1.29
1+:intercept_2016	–	−1.44	–	−1.26
1+:count_TRE	−0.07	−0.07	−0.07	−0.07
1+:friday	0.20	0.20	0.20	0.19
1+:full-time_student	−1.33	−1.33	−1.35	−1.28
1+:reside_in_PD1	−0.53	−0.53	−0.52	−0.49
1+:reside_in_PD2_6	−0.19	−0.19	−0.19	−0.18
1+:respondent	0.63	0.63	0.63	0.60
1+:sex_male	−0.18	−0.18	−0.18	−0.17
1+:suff_veh	0.44	0.45	0.45	0.42
1+:work_home_full-time	−0.41	−0.41	−0.41	−0.39
1+:work_outside_full-time	−0.90	−0.90	−0.91	−0.86
scale_2001	–	–	0.98	1.04
scale_2006	–	–	0.98	1.03
scale_2011	–	–	0.96	1.02
scale_2016	–	–	1.06	1.14
ρ^2 *against constants-only model*	0.080	0.080	0.080	0.080
ρ^2 *against null model*	0.405	0.405	0.405	0.405
Log-likelihood	−526,110	−525,940	−525,883	−525,856
AIC	1,052,250	1,051,902	1,051,795	1,051,750
Number of observations	1,275,214			

distribution are more useful for current purposes (refer to the supplementary material for a more detailed discussion and see Figures 5.1 and 5.2).

When the ρ^2 values are compared in binomial/trinomial logit models over the years, it can be seen that the 2016 models perform better than the models for other years. This can most likely be attributed to the availability, and hence use, of income data in that year, which is the major difference between the models.

Despite the low model performances of the models, most model parameters display a stable behaviour over the years. There is more fluctuation in the estimates when compared to the models of first work/school-related activity of the day, but again, this is expected. In addition, the fluctuations might also be attributed to having independent survey samples each year.

The results of the binomial and trinomial joint models are presented in Table 9 and Table 5.4 (in the supplementary material), respectively. Only the variables that are common to all the years are used in the joint models. It can be seen that the joint models produced close estimates to the year-specific models. Further, the estimates from four types of joint models are almost the same. This observation holds with or without scaling the estimates, as the year-specific scales are very close to 1.0. There is no clear pattern in terms of the changes in scales (in “JOSI” and “JO” models) or year-specific intercepts (in “NPDA” and "JO" models) over time. All joint models generate the same low goodness-of-fit, similar to the year-specific models, for presumably the same reasons.

Other

Due to the broad/heterogeneous definition of "other" activities, and inevitably, high variability in behaviour, models estimated for this activity purpose have very small explanatory power. Hence, detailed results are not shared here. Parameter stability follows similar trends described in marketing/shopping models. For the interested reader, associated tables and figures can be found in the supplementary material (Tables 6.1–6.6 and Figures 6.1 and 6.2) in addition to a discussion of the results. It is important to highlight that both respondent and online survey option have showed statistically significant impacts on the outcome.

Conclusions

Longitudinal behaviour of individuals might exhibit some changes or might remain stable depending on developments in regional land-use and transportation systems, changes in population attributes, and on the tastes and preferences of individuals, among other possible factors. While some things do not change—most employed individuals still have to go to work in the morning, students have to go to school, etc.—the impacts of improved technology (e.g., online teaching, online shopping, etc.) might be altering behaviour. This paper investigates the longitudinal stability of activity generation behaviour in the Greater Toronto and Hamilton Area (GTHA) for a 20-year period between 1996 and 2016 for various activity types (work-related activities, school-related activities, marketing/shopping and "other") using multiple model structures. Key findings are summarized below, first for work and school-related activities and then for non-work/school (NWS) activities.

Work/school-related activities

For both "work" and "school" purposes, the population is disaggregated into more uniform sub-samples considering the occupational categories/educational levels, and each activity purpose is subdivided into two considering their order in the day. Several valuable conclusions can be drawn by investigating the results of both year-specific and joint models developed for all the aforementioned subgroups adopting a utility-based approach:

First, the stable behaviour for first work/school activities exhibited by most parameters over the years and the fact that year-specific scales in the joint models are around 1.0 suggest that the year-to-year variations are low, and are all strong evidence that favour the use of such models for forecasting. That is, these models can be used to predict future activity generation behaviour confidently in the GTHA.

Second, since the year-specific models produce close estimates, joint models with different structures also yielded similar results with almost identical performance levels, and thus, one of the structures can be selected and used for forecasting without loss of information. In that case, the modeller faces a trade-off between the computational burden and the amount of detail desired in the model. Model building and estimation is less challenging for "NP" and "NPDA" types of joint models, whereas the amount of time and effort required increases if “JO” or “JOSI” types of joint models are estimated owing to the introduction of non-linearity into the problem by estimating year-specific scales, but in exchange additional information is obtained.

Third, the improving goodness-of-fit in *first work activity* models in time, when compared to the declining goodness-of-fit values in *first school activity* models in time raises questions regarding the sampling methods of the survey. Although the difference in the performance trends can be attributed to various other reasons (e.g., technological developments, online courses, etc.), one concern regarding the conduct of the survey is that the student population is not sampled as well as the worker population.

Fourth, there is a significant difference between the performances of the models for first work/school-related activity of the day and the models for second/subsequent work/school-related activity of the day, where the former ones obtain better fits. This is potentially an indicator that the first work/school-related activity of the day has a more systematic nature when compared to the subsequent activities of the same type, which are intrinsically more random, and thus, much more difficult to model capturing the underlying factors accurately.

Fifth, elementary and secondary school trips generally need only be modelled using average trip rates across the student population: parametric, utility-based models provide very little additional explanatory power.

Last but not least, the model results reveal that there is a significant respondent bias when the second/subsequent work/school-related activities of the day are considered, which suggests that the conduct of the survey might be improved by collecting information on daily schedules of individuals one by one, rather than by questioning a single household member who may not have complete information on all the other household members’ daily activities. An alternative remedy to this problem can be devised on the modelling side by tweaking the models to account for the known respondent bias.

Non-work/school-related activities

First, although most model parameters exhibit stable behaviour over the years and the year-specific scales are estimated to be close to 1.0—implying low year-to-year variations, developing robust parameterized models is challenging even when data from multiple surveys are exploited. Thus, it might be useful to test other modelling approaches, such as machine learning techniques, for modelling activity generation behaviour associated with NWS activities, which are more random in nature, to obtain models with higher fits.

Second, the model performances are very poor for both activity types: marketing/shopping and “other”, with marketing/shopping models having better fits. There are at least two reasons why the models for “other” activities perform worse than those estimated for marketing/shopping: (1) these activities are potentially more random and less systematically planned, when compared to shopping; and (2) the activity category is more heterogeneous than shopping, in that it covers all activities except work, school, and shopping, such as going to see a movie, visiting your friends, playing hockey, going to a dentist appointment, etc.

Third, several reasons are provided in Sect. 5 to explain the low fits of the models, but solutions have not been proposed. The problem can be mitigated to a certain extent, and thus, good-fitting models can be built if: (1) the activity definitions in the survey are improved, by adding more categories representing specific activity types; (2) data regarding activities/trips are collected for a longer time frame, instead of a single day; (3) a purpose imputation study is undertaken to match more specific purposes to NWS activities in Transportation Tomorrow Survey using comprehensive land-use data, which will facilitate the development of separate models for the identified purposes.

Fourth, the importance of the availability of income data, even aggregated into several categories at the household level, is evident when the performances of the year-specific models are compared, as there is a considerable improvement in the ρ^2 values of the models in the year 2016, in which income classes are used as variables. This observation holds for both activity types.

Fifth, survey design elements tested in this study were found to have crucial impact on activity/trip rates. For both marketing/shopping and "other" activities, there exists a significant survey method bias and respondent bias, hinting at considerable under-reporting of NWS trips in the region. This draws attention to the need of improving the conduct of the survey and/or incorporating the biases in the models developed to account for the aforementioned effects.

Despite the poor fit of the models, this study serves as a preliminary investigation that is conducted to have a better understanding of the general NWS activity/generation behaviour in the GTHA between 1996 and 2016 before building more advanced models.

All in all, the strong stability in model parameters over the years, where the variations remain in a very narrow band, is a significant conclusion considering that the standard assumption of travel demand forecasting is that the parameters remain constant over time, which is not tested except within the transferability literature. Based on our investigations, we find evidence supporting this assumption, encouraging forecasting applications, despite the developments in transportation-related aspects such as the availability of new modes (e.g., ride-hailing, bike-sharing, etc.) or the availability of real-time information (e.g., transit schedule information, etc.). What we would like to underline in this paper by "strong stability" is not that each and every parameter is consistently statistically equivalent in all the years, rather, we would like to emphasize that even though there are fluctuations—which are inevitable since the data come from independent samples in different time periods (each consecutive set is 5 years apart from the other) from a region that keeps developing (land use, transportation, economy, technology, etc.) (which might lead to changes in activity/travel-related behaviour) and growing with a very diverse population—the variations are small when their overall impact is considered during the period of analysis between 1996 and 2016. We believe that the stability observed (i.e., that the average rates, average propensities of attending a certain activity has not changed) in a 20 year period, which is a common time period for forecasting models, is a very useful information for activity-based travel demand modellers.

In the light of the results of this study, it can be concluded that taking advantage of multiple surveys and developing parameterized models are both useful for building policy-sensitive models. Since activity generation lies at the beginning of the decision-making process chain in an activity-based model, the outcomes of the component highly influence the remaining parts of the model, and hence, attaching importance to the internal processes of activity generation, such as frequency modelling as discussed in this paper, is critical.

A few limitations can be listed for this study. First, a distinction between separate activity locations has not been made in this study for work/school-related activities, but it is possible to distinguish the activities conducted at the usual workplaces or school locations and apply separate frequency generation models considering them as different activity purposes. Second, it is essential to disaggregate the population into more uniform groups, considering socio-demographics (e.g., age, employment status, occupation, etc.) of individuals, and build distinct models for each category for non-work/school purposes. Third, it may be useful to test the combined effects of survey design elements, estimating different bias parameters for respondent-phone, respondent-web, non-respondent-phone, non-respondent-web pairs. Nonetheless, this paper contributes to the travel demand modelling literature by investigating the activity generation frequency in the GTHA between 1996 and 2016 through the estimation of disaggregate econometric models which show strong stability over the years along with testing the impacts of survey design elements. Our future tasks cover developing more complex models while improving these models to incorporate more policy sensitivity with the goal of explaining the activity/travel behaviour better.

Acknowledgements This study was funded in part by an Ontario Graduate Scholarship and a Richard M. Soberman Fellowship (Department of Civil and Mineral Engineering, University of Toronto). Partial funding was also provided by a Natural Sciences and Engineering Research Council (Canada) Discovery Grant (RGPIN-2014-04479).

Authors' contribution GO: Analysis, literature review, manuscript writing, editing and content planning, EJM: Supervision on analysis and results, manuscript writing, editing and content planning.

Compliance with ethical standards

Conflict of interest This paper has no conflict of interest with any third party.

References

Ashby, B.: Transportation Tomorrow Survey 2016: data guide. Technical report. http://dmg.utoronto.ca/pdf/tts/2016/2016TTS_DataGuide.pdf (2018). Accessed 8 Aug 2018

Badoe, D.A.: An investigation into the long range transferability of work-trip discrete mode choice models. PhD dissertation, University of Toronto, Toronto, Canada (1994)

Badoe, D.A., Miller, E.J.: Comparison of alternative methods for updating disaggregate logit mode choice models. Transp. Res. Rec. **1493**, 90–100 (1995)

Badoe, D.A., Miller, E.J.: Modelling mode choice with data from two independent cross-sectional surveys: an investigation. Transp. Plan. Technol. **21**(4), 235–261 (1998). https://doi.org/10.1080/03081069808717611

Badoe, D.A., Steuart, G.N.: Impact of interviewing by proxy in travel survey conducted by telephone. J. Adv. Transp. **36**(1), 43–62 (2002)

Badoe, D.A., Steuart, G.N.: Urban and travel changes in the greater Toronto area and the transferability of trip generation models. Transp. Plan. Technol. **20**(4), 267–290 (1997). https://doi.org/10.1080/03081069708717594

Badoe, D.A., Wadhawan, B.: Jointly estimated cross-sectional mode choice models: specification and forecast performance. J. Transp. Eng. **128**(3), 259–269 (2002)

Caird, S., Lane, A., Swithenby, E.: Greening higher education qualification programmes with online learning. In: Leal Filho, W., Azeiteiro, U., Caeiro, S. (eds.) E-learning and Sustainability. Umweltbildung, Umweltkommunikation und Nachhaltigkeit—Environmental Education, Communication and Sustainability, vol. 35, pp. 105–116. Peter Lang Scientific Publishers, Frankfurt am Main (2014)

Cameron, A.C.: Advances in count data regression: I. Basic cross-section methods. In: 28th Annual Workshop in Applied Statistics, Southern California Chapter of the American Statistical Association. University of California, Los Angeles. http://cameron.econ.ucdavis.edu/racd/trcountI4up.pdf (2009). Accessed 7 Oct 2018

Cameron, A.C., Trivedi, P.K.: Regression Analysis of Count Data, 2nd edn. Cambridge University Press, Cambridge (2013)

Castiglione, J., Bradley, M., Gleibe, J.: Activity-Based Travel Demand Models: A Primer, Report S2-C46-RR-1. Transportation Research Board, Washington (2015)

Chung, B., Srikukenthiran, S., Habib, K.M.N., Miller, E.J.: The development of a web-survey builder (STAISI): designing household travel surveys for data accuracy and reduced response Burden. In: Presented at 11th International Conference on Transport Survey Methods. https://uttri.utoronto.ca/files/2017/09/-ISCTSC_11_paper_16-Development-of-a-Web-Survey-Builder-Platform-for-Household-Travel-Surveys.pdf (2017). Accessed 3 Mar 2019

Ciari, F., Marmolejo, A., Stahel, A., Axhausen, K.W.: Mobility patterns in Switzerland: past, present, future. In: Paper Presented at the 13th Swiss Transport Research Conference (STRC 2013), Ascona, Switzerland (2013)

Cotrus, A.V., Prashker, J.N., Shiftan, Y.: Spatial and temporal transferability of trip generation demand models in Israel. J. Transp. Stat. **8**(1), 37–56 (2005)

Croissant, Y.: mlogit: A Package for Estimation of Multinomial Logit Models in R (2019)

Data Management Group [University of Toronto]: 1996 Data Guide. http://dmg.utoronto.ca/pdf/tts/1996/dataguide.pdf (1997). Accessed 19 July 2018

Data Management Group [University of Toronto]: 2001 Data Guide. http://dmg.utoronto.ca/pdf/tts/2001/dataguide2001.pdf (2003). Accessed 19 July 2018

Data Management Group [University of Toronto]: 2006 Data Guide. http://dmg.utoronto.ca/pdf/tts/2006/dataguide2006_v1.pdf (2008). Accessed 19 July 2018

Data Management Group [University of Toronto]: 2011 Data Guide. http://dmg.utoronto.ca/pdf/tts/2011/dataguide2011.pdf (2013a). Accessed 19 July 2018

Data Management Group [University of Toronto]: 2011 TTS Version 1.0: Data Expansion and Validation. http://dmg.utoronto.ca/pdf/tts/2011/validation2011.pdf (2013b). Accessed 9 Sept 2018

Data Management Group [University of Toronto]: Reports. http://dmg.utoronto.ca/transportation-tomorrow-survey/tts-reports (2014). Accessed 5 Sept 2018

Dianat, L., Habib, K.M.N., Miller, E.J.: Modeling and forecasting daily non-work/school activity patterns in an activity-based model using skeleton schedule constraints. Transp. Res. Part A (2018)

Elias, W., Bekhor, S., Shiftan, Y.: Analysis of travel behavior in Arab communities in Israel: A Comparison Of Household Surveys. J. Transp. Geogr. **19**(1), 162–169 (2010). https://doi.org/10.1016/j.jtrangeo.2010.01.002

Farag, S., Schwanen, T., Dijst, M., Faber, J.: Shopping online and/or in-store? A structural equation model of the relationships between E-shopping and in-store shopping. Transp. Res. Part A **41**(2), 125–141 (2007). https://doi.org/10.1016/j.tra.2006.02.003

Ford, C.: Getting Started with Hurdle Models. University of Virginia Library: Research Data Services and Sciences. https://data.library.virginia.edu/getting-started-with-hurdle-models/ (2016). Accessed 25 Oct 2018

Fox, J.B.: Temporal transferability of mode-destination choice models. PhD dissertation, University of Leeds, Leeds, England (2015)

Golob, T.F., Regan, A.C.: Impacts of information technology on personal travel and commercial vehicle operations: research challenges and opportunities. Transp. Res. Part C **9**(2), 87–121 (2001)

Greene, W.H.: Econometric Analysis, 7th edn. Pearson Education Limited, Harlow (2011)

Habib, K.M.N.: Modelling activity generation processes. PhD Dissertation, University of Toronto, Toronto, Canada (2007)

Harding, C., Nasterska, M., Dianat, L., Miller, E.J.: Effect of land use and survey design on trip under-reporting in Montreal and Toronto's regional surveys. Eur. J. Transp. Infrastruct. Res. **18**(1), 36–59 (2018)

Hassounah, M.I., Cheah, L.-S., Steuart, G.N.: Under-reporting of trips in telephone interview travel surveys. Transp. Res. Rec. **1412**, 90–94 (1993)

Henningsen, A., Toomet, O.: maxLik: A package for maximum likelihood estimation in R. Comput. Stat. **26**(3), 443–458 (2011). https://doi.org/10.1007/s00180-010-0217-1

Hu, S.: Modelling trip generation/trip accessibility using logit models. PhD dissertation, Edinburgh Napier University, Edinburgh, Scotland (2010)

Huntsinger, L.F.: Temporal stability of trip generation models: an investigation of the role of model type and life cycle, area type, and accessibility variables. PhD dissertation, North Carolina State University, Raleigh, North Carolina, the United States of America (2012)

Inbakaran, C., Kroen, A.: Travel surveys: review of international survey methods. In: Prepared for 34th Australasian Transport Research Forum, Adelaide. https://www.researchgate.net/publication/-267365069_Travel_Surveys_-_Review_of_international_survey_methods (2011). Accessed 11 Mar 2019

Insee: Coefficient of Variation—CV—Definition. https://www.insee.fr/en/metadonnees/definition/c1366 (2016). Accessed 19 Jan 2019

Jackman, S.: pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory. United States Studies Centre, University of Sydney, Sydney, New South Wales, Australia, R package version 1.5.2. https://github.com/atahk/pscl/ (2017)

Kannel, E., Heathington, K.: The temporal stability of trip generation relationships. Technical paper (1972). https://doi.org/10.5703/1288284313837

Kleiber, C., Zeileis, A.: Visualizing count data regressions using rootograms. Am. Stat. (2016). https://doi.org/10.1080/00031305.2016.1173590

Lee, R.J., Sener, I.N., Mokhtarian, P.L., Handy, S.L.: Relationships between the online and in-store shopping frequency of Davis. Calif. Resid. Trans. Res. Part A **100**, 40–52 (2017)

McFadden, D.: Modeling the choice of residential location. In: Karlqvist, A., Snickars, F., Weibull, J. (eds.) Spatial Interaction Theory and Planning Models, pp. 75–96. North-Holland, Amsterdam (1978)

McFadden, D.: Structural analysis of discrete data with econometric applications. In: Manski, C.F., McFadden, D. (eds.) Econometric Models of Probabilistic Choice, pp. 198–272. M.I.T. Press, Cambridge (1981)

Miller, E.J.: Agent-based activity/travel microsimulation: what's next? In: Briassoulis, H., Kavroudakis, D., Soulakellis, N. (eds.) The Practice of Spatial Analysis: Essays in Memory of Professor Pavlos Kanaroglou, pp. 119–150. Springer, Berlin (2018)

Miller, E.J., Roorda, M.J.: A Prototype model of household activity/travel scheduling. Transp. Res. Rec. J. Transp. Res. Board **1831**, 114–121 (2003). https://doi.org/10.3141/1831-13

Miller, E.J., Dalton, P., Briggs, R.: GTA trip generation rates, 1986–1996. Technical report. University of Toronto (1998)

Miller, E.J., Roorda, M.J., Carrasco, J.A.: A tour-based model of travel mode choice. Transportation **32**, 399–422 (2005). https://doi.org/10.1007/s11116-004-7962-3

Montgomery, D.C., Runger, G.C.: Applied Statistics and Probability for Engineers. Wiley, New York (2003)

Mullahy, J.: Specification and testing of some modified count data models. J. Econom. **33**, 341–365 (1986)

Mwakalonge, J.L., Badoe, D.A.: Trip generation modeling using data collected in single and repeated cross-sectional surveys. J. Adv. Transp. **48**(4), 318–331 (2012). https://doi.org/10.1002/atr.217

de Dios Ortúzar, J., Willumsen, L.G.: Modelling Transport, 4th edn. Wiley, Chichester (2001)

Ozonder, G.: Technical documentation on cleansing transportation tomorrow survey data. Technical report, University of Toronto (2018)

Richardson, A.J.: Behavioural mechanism of non-response in mail-back travel surveys. Transp. Res. Rec. **1855**, 191–199 (2003)

Richardson, A.J.: Sample bias in telephone interview travel surveys. In: Prepared for 10th Australian transport research forum, Melbourne, Australia. https://atrf.info/papers/1985/1985_Richardson.pdf (1985). Accessed 26 Feb 2018

Roorda, M.J., Morency, C., Woo, K.: Two cities, two realities? Trans. Res. Rec. J. Transp. Res. Board **2082**(1), 156–167 (2008). https://doi.org/10.3141/2082-19

Rose, A.: Transportation Tomorrow Survey 2016: data expansion and validation. Technical report. http://dmg.utoronto.ca/pdf/tts/2016/2016TTS_DataExpansion.pdf (2018). Accessed 14 Nov 2018

Rotem-Mindali, O., Weltevreden, J.W.J.: Transport effects of e-commerce: what can be learned after years of research. Transportation **40**, 867–885 (2013). https://doi.org/10.1007/s11116-013-9457-6

Roy, R., Potter, S., Yarrow, K.: Designing low carbon higher education systems: environmental impacts of campus and distance learning systems. Int. J. Sustain. High. Educ. **9**(2), 116–130 (2008)

RStudio Team [Computer Software]. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com/ (2019)

Russell, J.N., Bose, J., Giesbrecht, L.H.: Nonresponse bias in a travel survey of nontelephone households. In: Prepared for 2004 AAPOR Annual Meetings. https://nces.ed.gov/FCSM/pdf/IHSNG_AAPOR04_NR_FinalPaper.pdf (2004). Accessed 23 Dec 2018

Salem, S.M.K.M., Habib, K.M.N.: Use of repeated cross-sectional travel surveys to develop a meta model of activity-travel generation process models: accounting for changing preference in time expenditure choices. Transp. A Transp. Sci. **11**(8), 729–749 (2015). https://doi.org/10.1080/23249935.2015.1066900

Salem, S.M.K.M.: Temporal transferability of model components within an activity-based travel demand modelling approach. PhD dissertation, University of Toronto, Toronto, Canada (2016)

Shams, K., Jin, X., Argote, J.: Examining temporal transferability of trip frequency choice models. In: Transportation research board 93rd annual meeting, Washington, DC (2014)

Stopher, P.R., Jones, P.M.: Developing standards of transport survey quality. Keynote paper prepared for the international conference on transport survey quality and innovation: how to recognize it and how to achieve it, Kruger Park, South Africa. https://www.researchgate.net/publication/267242282-_developing_standards_of_transport_survey_quality (2001). Accessed 21 Feb 2018

Suel, E., Polak, J.W.: Incorporating online shopping into travel demand modelling: challenges. Prog. Oppor. Transp. Rev. **35**(5), 576–601 (2018). https://doi.org/10.1080/01441647.2017.1381864

Susmel, R.: Lecture 7: count data models. Lecture notes. https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf (2015). Accessed 4 Feb 2019

Train, K.: Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge (2003)

Travel Modelling Group [University of Toronto]: GTAModel V4.0: design, background, and calibration. http://tmg.utoronto.ca/documents/ (2015). Accessed 18 Oct 2017

Versteijlen, M., Salgado, F.P., Groesbeek, M.J., Counotte, A.: Pros and cons of online education as a measure to reduce carbon emissions in higher education in the Netherlands. Curr. Opin. Environ. Sustain. **28**, 80–89 (2017)

Weltevreden, W.J.W., Van Rietbergen, T.: E-shopping versus city centre shopping: the role of perceived city centre attractiveness. Role Perceived City Centre Attract. **98**(1), 68–85 (2007). https://doi.org/10.1111/j.1467-9663.2007.00377.x

Winslott Hiselius, L., Rosqvist, L.S., Adell, E.: Travel behaviour of online shoppers in Sweden. Transp. Telecommun. **16**(1), 21–30 (2015). https://doi.org/10.1515/ttj-2015-0003

Yasmin, F., Morency, C., Roorda, M.J.: Trend analysis of activity generation attributes over time. Transportation **44**(1), 69–89 (2017). https://doi.org/10.1007/s11116-015-9624-z

Yunker, K.R.: Tests of the temporal stability of travel simulation models in southeastern Wisconsin. Transp. Res. Rec. **610**, 1–5 (1976)

Zeileis, A., Kleiber, C., Jackman, S.: Regression models for count data in R. J. Stat. Softw. **27**, 1 (2008). https://doi.org/10.18637/jss.v027.i08

Gozde Ozonder has B.Sc. (Civil Engineering) and M.Sc. (Engineering Management) degrees from the Middle East Technical University. She is currently a Ph.D. Candidate at the University of Toronto. Her research focusses on redesigning and reconstructing the components of the agent-based activity-based travel demand model TASHA that is commonly used in the Greater Toronto and Hamilton Area.

Dr. Eric J. Miller (Professor) has been a faculty member in the Department of Civil and Mineral Engineering, University of Toronto since 1983, where he is currently Director of the University of Toronto Transportation Research Institute. He has BASc. and MASc. degrees from the University of Toronto and a Ph.D. from MIT. He is recipient of the 2018 International Association for Travel Behaviour Research Lifetime Achievement Award.