ECON 6805 – Transportation Economics – Justin Desrosier

Passenger Rail Demand Estimation: A Framework for the Windsor – Quebec City Corridor

Contents

Introduction

The Windsor – Quebec City corridor is a geographic term for the area that spans the most populous region of Canada, encompassing 2 of its 3 largest metropolitan areas, and the nation's capital. Passenger Rail in the corridor directly connects 18 census metropolitan areas with populations greater than 100,000 and in combination with indirect connections from rural regions without stations, the rail line provides service to 18 million people in the combined CMAs and near rural regions. The railway in the corridor is highly developed for both passenger travel services and rail freight transportation, which operate on the same infrastructure. Intercity passenger traffic in the region is also served by modern highways and airports.

Passenger rail service in the corridor has been provided by Via Rail since their 1977 formation following a split-off from CN Railway, the larger of the two major rail carriers in Canada. Via Rail is a crown corporation that is partially owned by the federal government. The policy decision of the government to support Via Rail was made with an objective of offsetting costs of providing passenger rail service during a time of increasing personal vehicle and air travel (Nelligan, 1982). Since any firm providing passenger rail service at the time would operate with negative profits, the government decided to subsidize a firm (Via Rail) to provide business incentives and cater to the public interest in enabling connection of remote communities.

There is significant interest in developing models to estimate the current and projected demand for passenger rail travel in the corridor. Upgrades to the service capacity (frequency) and the service quality/speed have been underway or in a planning phase for the lifespan of the rail carrier. There have also been numerous feasibility reports and studies published by different governments dating back to at least the 1990s concerning high-speed rail in the corridor (SNC-Lavlin et al., 1995; Transport Canada et al., 1995).

Rough estimates using Statistics Canada population projections and a census analysis of corridor residents would indicate that 23.2 million people will be living in the near service area by the year 2043 (Statistics Canada, 2019; Statistics Canada, 2006). In anticipation of a potential wave of future demand in the corridor, econometric models are a pertinent method to decrease uncertainty about an investment in rail transportation fixed facilities. By uncovering the causal factors that drive the current rail travel demand, a researcher could develop evidence-based results to support the predicted future demand when there are changing demographic factors. Therefore, the objective of this research paper is to first review the literature on the established modelling frameworks developed for passenger transportation demand estimation, secondly, the paper will provide a recommendation for the most applicable modelling practice in the context of the corridor's transportation network characteristics. The third portion of this paper will detail the analytical specifications and data requirements of gravity models and discrete choice models as the recommended methods to estimate travel demand in this setting. Final conclusions for the paper will detail the expected results based on previous findings, and the limitations facing this approach.

It should be noted that the research results from this study for the most part could be considered a steppingstone for further quantitative research. The results section will be necessarily simplistic as no technical research was carried out in this research. Instead, from an interpretation of past research on the topic, there will be included potential results subject to meeting the data requirements mentioned in the methodological section.

Literature Review

There have been many attempts at estimating travel demand using models that are specific to the mode of interest, consumer market, and the provider of service. Ad hoc models exploit the natural supply and demand relationship between the price of service (fare) and quantity of service but lack the ability to describe passenger behaviour at a granular level. The gravity model is an example of an ad-hoc passenger demand model, as it leverages the underlying relationship between passenger transportation volume between an origin and destination as a function of the number of potential travellers (population) and their propensity to travel as a function of GDP among other variables, while factoring distance as a proxy for transportation cost. However, the gravity model in its functional form does not provide sufficient behavioural insight into an individual consumer's travel choice. The gravity model is best used for aggregate estimates of passenger transportation demand between two economic regions. Traffic engineering models can also provide insight into the network effects pertaining to travel decisions. These models, however, are most suitable in an urban context where the demand for transportation services (especially public transportation) is almost entirely derived from a consumer's need to travel somewhere in order to participate in the economy (go to school, work, go to the mall). Traffic engineering models in this context are geared toward analysis of all nodes and transportation options in a network and minimizing the time required to travel.

The contemporary literature on the topic of discrete choice travel modelling has a foundation on the concept of utility maximization with respect to a money budget constraint and often a time budget constraint. For instance, among many other transportation researchers, Johnson (1966) provides the early literature with a quintessential conceptualization of this technique. In his work on travel time and the price of leisure, he illuminates the importance of time scarcity when estimating demand for travel given the opportunity costs associated with using that time, approximated by the hourly wage rate. His paper sought to "capture the essence of decisions pertaining to trips, work, leisure, and other competing uses of time" and solve a utility maximization problem to determine the price a consumer is willing to pay for travel.

It is clear that traffic engineering models are not suitable for intercity passenger rail travel estimation in the Windsor-Quebec City corridor, the primary objective of the model is not consistent with the problem of demand for travel given regional economic factors, or the problem of modal choice given individual preferences and budget allocation. Traffic engineering models could provide insight on modal choice due to service efficiency, but the DCM and gravity approach will cover this consideration sufficiently. Indeed, in order to estimate the demand for rail passenger service in the corridor, it will be essential to test the elasticity with respect to price, travel time, and socioeconomic characteristics for each of the alternative intercity transportation modes, car, bus, rail, air, and control for specific elements of each parameter as will be discussed further. While a traffic engineering model may provide insight on optimizing for the minimum travel time of a mode or combination of modes, it cannot provide us with a means to test these elasticities between modes given revealed or stated preference data. This is especially a problem for intercity rail travel demand where a key factor will be consumers propensity to travel via a specific mode. Furthermore, in the case of travel along this corridor; it is commonly for leisure purposes and therefore preference for a given transport mode may not be derived solely from its ability to transport the passenger to a destination (Mokhtarian et al., 2001). For instance, car travel may be slower than rail but ultimately preferred due to the freedom to choose alternate roots or stops, a traffic engineering model would be incapable of capturing this variation in preference, instead favouring the optimal mode given price.

An ad-hoc model is a good candidate for intercity rail travel demand estimation, especially the gravity model. The method of OLS, with specified explanatory variables related to the number of travellers between each of the corridor's city pairs, could provide robust estimates of the number of potential travellers in the entire corridor for a given month, as well as forecasts (Jones & Nichols, 1983; Owen & Phillips, 1987; Anderson, 2011). However, an ad-hoc model would be too broad in its differentiation of traveller's demand for specific modes, similar to the traffic engineering approach. As mentioned, this is a crucial aspect when it comes to determining intercity passenger travel demand for rail. The ad-hoc model falls short in this respect but may prove a strong addition to an analysis of the aggregate demand for travel when used in conjunction with a discrete choice model.

Discrete choice models prove to be a strong candidate for modelling passenger travel demand when it comes to research interest in a specific mode against competing alternatives. Maximum likelihood estimation of a discrete set of transportation choices is a powerful toolkit for the researcher interested in estimating the elasticities associated with travel choices when a consumer is faced with several viable options. For example, In Johnson's (1966) model of the demand for travel, the consumers' equilibrium is found using two budget constraints

$$(1)\ P_x X + P_c C = P_w W$$

$$(2)\ T_x X + W + L = T_o$$

Where X, C, and W are respectively the number of pleasure trips, units of the non-trip commodity, and hours of work. P represents the price of the respective subscript for each parameter. Making $P_w W$ the total income over a given period, assumed to be entirely spent in the case of a utility-maximizing combination of trips and non-trip commodities. T_x is a parameter representing the time price of trips, L represents the hours of leisure, and T_o denotes the fixed time parameter which is the total time duration of a given period. The model assumes that T_o is entirely allocated between the potential choices of X, W, and L, as the non-trip commodity does not require time to consume for simplicity of the model. Finally, the usual assumption applies that a consumer's preference for each of the four consumption items, X, C, W, and L can be represented by a utility function. In equilibrium, the marginal utility of the composite commodity is equal to the marginal utility of income times the price paid for the commodity. Johnson exemplifies the early thinking on the topic of time constraints in the utility maximization problem in the context of estimating the determinants of passenger transportation demand. Along with his contemporaries, he set the stage for more complicated activity-based models that include parameters specific to transportation mode choices, such as service level quality and travel speed. Several discrete choice travel demand projects followed in the direction of Johnson's footsteps and led to further development of the time-constrained utility maximization problem in transportation.

Ben-Akiva and Lerman (1985) wrote a book on discrete choice analysis with applications to travel demand. They state the interest of discrete choice models as modelling choice from a set of mutually exclusive and collectively exhaustive alternatives with the objective of maximizing utility subject to specific constraints on consumption. In other words, and in terms of passenger transportation demand, a utility function is an expression of a consumer's preferred combination of goods out of all potential goods that they could consume, subject to their total money budget, and often time budget, that they have available. In practice, the utility of the alternative consumption goods are considered random variables, so the probability that an alternative is chosen is defined as the probability that it has the greatest utility among the available alternatives.

In terms of what variables to include when considering the parameters that affect the travel preferences of the population living in the Windsor-Quebec City corridor, Dargay and Clark (2012) provide a valuable collection of recent thoughts on the topic. Factors affecting the parameters required for modelling the demand for rail travel between the major stations along the corridor have extensively been covered by these authors. In their paper on the determinants of long-distance travel in Great Britain, a similar country in terms of per capita GDP, the authors synthesize several elasticity findings from previous studies that contribute to the research on demand for travel given socio-economic attributes of a population in consideration. Dargay and Clark then go on to estimate coefficients of a linear regression equation specified for each travel mode, travel purpose, journey length, and a number of mode-purpose-distance combinations using the National Travel Survey in the UK. Demand is defined as person miles and is correlated with travel costs, travel time, and socioeconomic and demographic characteristics of a population by a set of elasticities. Substitution by mode is captured through cross-elasticities for travel costs and time. All elasticities varied by the purpose of travel, as well as distance and mode. The author's findings are summarized in Table 1 to highlight the several explanatory variables they used as determinants to long-distance travel, as well as their findings from previous work on the topic.

Table 1 – Dargay and Clark (2012) Findings

Reference Article	Explanatory Variables	Dependent Variable(s)	Data	Findings	Results from Dargay and Clark (2012)
Mokhtarian et al. (2001) – OLS regression models with a central tenet that individuals may perceive positive utility in the journey itself, and the inclusion of explanatory variables that reflect this.	Traditional demographic factors, variables that captured objective mobility (frequency of trips by various purposes), measure of excess travel (tendency toward an unnecessary journey)	Total long-distance travel (over 100 miles) by vehicle or airplane controlled for work or leisure.	Be-spoke travel survey reported by individuals in San Francisco	Purposeful journeys were seen to have a positive relationship with distance travelled, while unnecessary journeys, the opposite. Attitude, personality, lifestyle may contribute to distance travelled.	
Georgii and Pendyala (2001) and Mallett (2001) – grouping by age bands, and income measures, the authors fit OLS regression equations for each group.	Income, age, household structure, ethnicity, education, employment, car availability.	Rate of travelling (number of trips in a given time period of 100 miles or more)	1995 American Travel Survey	High income households made four times as many trips annually as those in the lowest income group, with the smaller average from low-income group due to 60-80% of people not making any trip during the year. The most immobile group were those in households without a vehicle. Car was the dominant mode for all income groups, but low income were more car dependent.	Likelihood of long-distance travel was considerably lower for those without car availability. Income was found to be highly significant and had an elasticity of 0.51 at mean distance and income.
Limtanakool et al. (2006a) & (2006b) – series of binary logit models	Gender, age, income, household composition, car availability, day type (control for holidays), population density - (2006b) added measures of land use density at destination and a ratio of rail v.s. car journey time.	Likelihood of making a long-distance journey of a particular type (business/lesiure) using a particular mode (train/car)	UK & Netherlands National Travel Survey	Being female, living in more complex households (number of residents), and density of city of residence reduced the probability of making a long-distance trip. High income increased the probability.	Time trend was insignificant, there was no trend in travel over time not explained by other explanatory variables. Long distance travel was lower for women. Long distance travel was found to decrease as household size increases.
Orfeuil and Soleyret (2002) – using univariate and cross-tabulations	Age, family size, income, degree of urbanisation (conurbation size and proximity to urban centre)	Average distance travelled per week	1994 French National Transport Survey	Long distance travel was made more often by middle aged groups and higher incomes. Amount of long-distance travel was higher for those living in central areas, and in this regard, less by car.	Long distance travel was greater for those under 60, those employed, and students. Long distance travel increased as the population of a conurbation decreases and is greatest for those living in rural areas.

Bradley (1988), raises an important concern when considering the external validity of stated preference data as a function of the realism and relevance of hypothetical travel choices. Particularly, he draws light on factors affecting travel demand estimation, such as time constraints, alternative mode constraints, and linkages between individuals. These factors tend to limit or influence the observed choices people make in actual contexts, he argues and therefore may cause spurious results if data from hypothetical situations are used. In the context of intercity rail passenger demand estimation, linkages between individuals as mentioned by Bradley may be a difficult effect to capture when modelling the discrete choices between travel modes in the Windsor-Quebec City corridor. In particular, travellers' preference for an intercity transportation mode that does not require another person to pick them up in their destination city as a link to their final destination (last-mile) may produce a different hypothetical result than in an actual context. As alluded by Wardman (2007) in his cross-sectional analysis of access and egress times at stations, this problem could be partially mitigated by including a weight to indicate the level of public transportation connectivity that a destination city has in order to connect travellers from the airport, train, or bus terminal to their home (final destination).

Table 2 summarizes the findings of the early literature on the subject of passenger travel demand with a focus on rail transportation and illustrates the differences between variables and technical specifications of the ad-hoc and discrete choice modelling approaches.

Modelling Methodology

From previous research on the topic of passenger rail travel demand and general discussion of travel demand estimation in transportation economics, it is most logical to leverage discrete choice modelling for an investigation of the Windsor – Quebec City corridor. The evidence to support this modelling framework follows from the remarks found in the literature review. In addition, a comprehensive study of passenger rail travel in the corridor could benefit from ad-hoc supply and demand analysis using gravity model specifications. Generating results from the two approaches can provide the researcher with a more complete report employing the advantages of each method. Generally, the advantage of DCM is robust elasticities describing willingness to travel for a given modal choice and passenger-specific characteristics, such as income, or travel time/service quality. For the ad-hoc gravity model, regression coefficients could provide causal insight on passenger flows between population centres while considering regional socio-economic determinants. This

Table 2 – Railway Passenger Transportation Demand Estimation

	Explanatory Variables	Dependent	Main Results	Methodology
Ad Hoc				
Owen and Philips (1987)	GDP, revenue per journey, time trend (car usage, economic and population), high speed train (HST) (service quality), coach and air competition	Number of journeys between two stations	Distinguishes between short- and long-term responses to changing economic conditions. Short term fare elasticities with respect to total journeys were lower than long-term fare elasticities (-0.69, -1.08) and for GDP (0.93, 1.39). Coach competition had an elasticity of -0.07, -0.08 in the short- and long-term. Introduction of HST had a positive effect on patronage 0.17, 0.23 short- and long-term elasticities.	A single equation (OLS) model for each origin-destination rail service stemming from London, UK to other
Jones and Nichols (1983)	GDP, revenue per journey, rail departure frequency, rail journey time, cyclical fluctuation, fuel costs, quality level of non-rail service, seasonal dummy	Ticket sales from a London, UK origin station to a provincial destination station.	Demand is significantly affected by the average fare paid, retail price inflation, rail journey time, level of competing modal services, cyclical economic activity, and seasonality. No evidence of a response to changes in train departure frequency or the level of petrol prices (fuel costs). Profit effects of higher speed trains are uncertain without estimates of the capital and operating costs.	A single equation (OLS) model for each origin-destination rail service stemming from London, UK to other regions.
Behavioural				
Oum and Gillen (1983)	Income and leisure-time constraints (price and quantities of rail, bus, & air travel, other goods & services, and the wage rate opportunity cost of time spent travelling)	Demand for travel as derived from the overall utility maximizing behaviour of an individual with respect to all transportation and other goods choices.	Results indicated that consumer demand preferences are time- and season- varying. Average weekly work hours were found independent of travel demand – indicating possibly a poor variable for approximation of a leisure time constraint, or an outcome of unsegmented business and non-business travel data. Non-homothetic, non-separable preference structure was found to be a favourable estimation strategy – indicating the demand system for intercity passenger modes is inextricably linked to the rest of the economy.	Direct demand function derivation using a translog reciprocal indirect utility function, for each of 5 consumption goods (air, bus, rail, other goods, other services) (i.e. nonlinear least squares equivalent to maximum likelihood method)
McFadden (1974)	Family Income, car-bus cost per round trip, bus walk time times the hourly wage rate, bus wait time times the hourly wage rate, bus travel time times the wage rate, our auto mode preference effect.	Auto-bus mode choice.	Likelihood ratio index was a more stable and statistically satisfactory measure for estimation. The authors were able to estimate robust figures for the hourly value of on-vehicle time, and wait time. They were unable to reject the null hypothesis that all components of travel time are weighed equally. The value of time may be an increasing function of the wage rate, consistent with hours-worked decisions closely approximating the neoclassical labour-leisure margin.	Using 213 households living and working in the San Francisco Bay area. Binary logit model estimated by maximum likelihood methods. Note that specifically McFadden is looking at work trip behaviour, since this model is of urban rail travel.

section will explore further the methodology and specifications of each model and provide a recommendation to the quantitative analyst on the technique to utilize for a picture of the demand for passenger rail travel in the corridor.

There are primarily two problems facing the researcher that is interested in estimating the demand for passenger travel by rail in most contexts, but certainly for the case of the Windsor – Quebec City corridor. There is first the question of how many people are expected to travel between each node (population centre) for the duration of a given time period, and it follows that there is the question of what percentage will choose to travel by rail or any other of the alternative transportation modes for that matter. For a researcher to determine the number of people expected to travel between each node in the corridor, gravity modelling should first be developed. Once a researcher has undergone a gravity model analysis of passenger network flows, discrete choice modelling should follow as a better approach to study the determinants of modal choice in the corridor. Both models would ideally leverage many years of recent historical data and could be used in tandem with population growth and other demand-related projections to predict forthcoming rail passenger volume if the assumptions concerning causality are deemed accurate.

The Gravity Model

The gravity model approach and its underlying assumptions are borrowed from Newton's law of universal gravitation. The leading use of classical gravity equations in economics is in international trade, where the difference of GDP over the distance between two countries has empirically been shown to be indicative of their trade relationship akin to the orbital velocity between two celestial objects (Anderson, 2011). In the context of passenger demand for rail travel in the Windsor – Quebec City corridor, results from this model could provide supporting evidence on the causal effects that influence the differing volume of travellers between population centres. In this regard, the effect of GDP and population, or growth, on the volume of travellers could be estimated for each node in the corridor.

For this portion of the analysis, the ideal baseline data that would be required for each of the 18 CMAs (nodes) would be regional GDP, population, and distance between each of the 18 CMAs for a total of 306 network pairs (N*(N-1), where N=# of CMAs). This data would form the independent variables of a regression equation. Data on the number of travellers between each CMA would form the dependent variable. For the rail mode, travel volume between each CMA could be as granular as daily ticket sales, but the time frame of the panel data would be subject to time frame consistency across the available data, often GDP is only released quarterly and for regional GDP, annually so a yearly analysis may be the limit. The theoretical gravity equation, in this case, would take the form:

$$(3) \quad V_{t,CMA_i,CMA_j} = \frac{\left(GDP_{t,CMA_i}/Population_{t,CMA_i}\right) * \left(GDP_{t,CMA_j}/Population_{t,CMA_j}\right)}{Distance_{CMA_i,CMA_j}}$$

Where the dependent variable, V_{CMA_i,CMA_j} would represent the volume of travellers, as total number of individual journeys (ticket sales or highway camera data) between CMA_i and CMA_j in time-period t. The econometric specification as commonly expressed in logarithmic form could then be expressed as:

$$\begin{aligned}(4) \quad & log(V_{t,CMA_i,CMA_j}) \\ & = \beta_1 log(GDP_{t,CMA_i}) + \beta_2 log(GDP_{t,CMA_j}) + \beta_3 log(Population_{t,CMA_i}) \\ & + \beta_4 log(Population_{t,CMA_j}) - \beta_5 Distance_{CMA_i} - \beta_6 Distance_{CMA_j} \\ & + \beta_7 X_{t,CMA_i,CMA_j} + \lambda_{t,CMA_i,CMA_j} + \theta_{CMA_i,CMA_j} + \varepsilon_{t,CMA_i,CMA_j}\end{aligned}$$

Where X_{t,CMA_i,CMA_j} could be any other variables, where data is available, that reflect the determinants to passenger travel between CMA_i and CMA_j such as the price of a train ticket specific to travel between the 2 CMAs and time period t, or perhaps the attractiveness of a CMA for tourism. λ_{t,CMA_i,CMA_j} would represent random effects for travel during the time-period t between CMA_i and CMA_j, that are not captured in the data, such as an unexpected closure of the railway due to environmental causes. θ_{t,CMA_i,CMA_j} would represent time-fixed effects between CMA_i and CMA_j that are not captured in the data, for example, demand for travel could have seasonal components (Jones & Nichols 1983), whether the research would choose a fixed or random effects model would follow the indication of a Hausman test. $\varepsilon_{t,CMA_i,CMA_j}$ would be the error term capturing variation in the dependent variable that could not be explained by available data, for example, it is possible, that a strong desire of travellers to take a train route through a scenic portion of the corridor may drive rail passenger demand, whereas another route may be considered less scenic and therefore passengers may be, however slightly, more inclined to purchase a faster travel option such as a flight.

The results from this regression specification using gravity model theory have a drawback in the context of travel in the Windsor – Quebec City corridor that stems from the nature of a network with many nodes. If a panel data set was to be compiled with data for each of the 306 possible origin-destination CMA pairs in the network, the model may be limited in its scope of multi-node journeys and there would be underestimated demand for travel between two CMAs. The transportation researcher's most applicable dataset for this dependent variable would include on the ticket sales data, all the nodes that each passenger travelled through, thus allowing more accurate network flow information between each of the nodes. Consider the portion of the corridor between Toronto and Oshawa as an example, in this case, Oshawa is the nearest eastward CMA (although some consider it a part of the GTA) and all the travellers that are due for a destination further east along the corridor would also travel through Oshawa, furthermore, anyone travelling from an origin east of Oshawa to a destination west of Oshawa would have to pass through. Therefore, the gravity model estimation of travel demand would underestimate the total travel volume between Toronto and Oshawa significantly if a researcher were to observe the coefficient for this specific V_{t,CMA_i,CMA_j}. Therefore, the centrality of a CMA in the corridor, with a weighting of the total population of all eastward CMAs and westward for that matter, would be an omitted variable in this framework. In this regard, it would be necessary to apply some post regression calculations for a picture of the total volume between each neighbouring CMA pair in the network.

To predict the increase in passenger volume between each neighbouring CMA pair in the corridor, it would first be necessary to have the regional population and GDP growth forecasts. With this information, a researcher could find the unique coefficients using equation (1) specified for each of the 306 (18*(18-1)) CMA origin-destinations and their reverse directions. For example, if the natural logarithm of GDP_{t,CMA_i} was found to be 0.2, that would indicate that a 1% increase in the regional GDP of CMA_i during the next time-period t+1 would increase passenger volume by 0.2%. Therefore, if GDP growth in CMA_i was projected to increase by 5% in 5 years, it could be said, using GDP growth as a scalar on the coefficient GDP_{t,CMA_i}, what the predicted passenger travel increase would be. In the mentioned example of a GDP_{t,CMA_i} being 0.2, and a 5 year regional GDP growth projection of 5%, we would have a 5% X 0.2 = 1% predicted increase in passenger volume between Windsor and London over 5 years due to the increase in regional GDP (assuming that passenger volume is linearly increasing), the same could be applied for population growth and other parameters. The analyst could then sum across all journeys that overlap a given neighbouring CMA pair to find the total percentage increase in passenger volume predicted for any of the 18 CMA pairs.

The Discrete Choice Model

Now we have estimated the causal factors affecting the total number of people expected to travel on all modes during any given time period and given the projections of our independent variable economic data, we can make a prediction about future demand. However, the gravity model alone is not a comprehensive approach to estimate the demand for any specific transportation mode. For a more thorough analysis of the determinants of travel demand using a specific transportation mode, quantitative research should include a discrete choice model estimated using the maximum likelihood method with a multinomial logit specification. The DCM of modal choice for travel between the CMAs in the Windsor – Quebec City corridor encompasses the factors affecting individual preferences for a certain mode of travel as well as characteristics of each mode and therefore will allow the researcher to estimate the change between modes given an anticipated or projected change in consumer preference structure. Since in the case of the gravity model, we take the dependent variable of ticket sales or aggregate travel data as a given, the estimates would assume no change in consumer preferences for the choice of travel mode over time. A discrete choice approach will allow the researcher to employ utility theory for a more in-depth look at the causal factors affecting the proportion of the population that decides to travel by rail.

The data required for the maximum likelihood specification of a multinomial logit model will fall into two categories, data describing characteristics of each transportation mode, and data describing the characteristics of each person who either travelled or did not travel on a given mode type. The data describing each modal choice should be homogenous across modes, ideally, there should be an objective measure of service quality accounting for factors such as comfort, on-board meals, internet access, ability to sleep, etc., as well there should be data on travel speeds (Johnson 1966). Ultimately, we would like to construct a measure from the data on modal characteristics to create an index that captures the Value of Travel Time (VOTT) for each mode in this analysis (Correira et al., 2019). There could also be objective data on the ease of access to a station/terminal in each CMA, or efficiency of destination station/terminal to the final destination (last-mile) of travel in each CMA measured possibly by public transportation availability/cost (McFadden, 1974; Bradley, 1988; Wardman, 2007; Wong et al., 2015).

In terms of the data required for each individual, there are many factors that may affect their decision to travel and their choice of mode. Socio-economic data could first be considered such as income/wage rate opportunity cost, age, household structure, education (Georgii & Pendyala, 2001; Mallett, 2001). Other individual-level characteristics could also be considered, such as the reason for travel (business/leisure), availability of a car, and population density of their census subdivision for example (Mokhtarian et al., 2001; Limtanakool, 2006a, 2006b; Orfeuil & Soleyret 2002). The data for individual characteristics should also include CMA level data, such as the efficiency of public transportation networks used for station/terminal to final destination (last-mile) portion of the trip, but also regional GDP, airport/terminal/station quality, primary industry, distance to nearest CMA, size of nearest CMA all may have an effect on residents travel mode choice for each CMA. In addition, we also require a measure of the ratio of each individual's income spent on all non-transportation goods vs. intercity transportation as well as the same ratio for time spent consuming these goods to gauge the value of travel time (VOTT) among individuals. This data will be needed when formulating budget constraints, the VOTT has often been considered the opportunity cost given by wage rates (Becker, 1965; Johnson, 1966; Jara-Diaz, 2000). Lastly, it is necessary to have a taste parameter for each individual, this could be survey data on how much each individual prefers comfort during travel or freedom to alter their route (which is the case of private vehicles), alternatively, this taste parameter could be derived from the other socio-economic data mentioned.

The dependent variable would necessarily be the same data on ticket sales, or survey travel data as the gravity model, that is, the aggregate number of travellers by each mode for all time periods between each of the 306 pairs of 18 CMAs in both directions. A note should be made regarding the business/leisure binary variable for each trip that was mentioned, it is a crucial element of the equation when it comes to the derived demand nature of transportation consumption. Since the use of travel is considered to be for an expressed activity at the traveller's destination, and not for the sake of travel itself, a business/leisure binary variable should account for some of the variations in the enjoyment of travel, particularly when it is phrased in an example such as the difference between travellers value of travel time for a private vehicle traveller that enjoys taking the scenic route under no time constraints (leisure) vs. the traveller (business) who needs to leave in the morning and be back in the afternoon. This consideration is elaborated on later in this DCM methodology section with insight from Correia et al. (2019).

When compiling the panel dataset with these variables the researcher would want to be sure that CMA boundaries, the number of years, the dependent variable aggregate number of trips, and any other economic data overlapping from the gravity model are the same for the consistency of both estimates. It would also be ideal to have a unique traveller ID linked to all individuals with their demographic/socio-economic characteristics and their travel choices, that being choice of mode and frequency of trips or total distance travelled between CMAs. An excellent data set would be a longitudinal cohort of at least 5000 randomly selected individual travellers (and non-travellers) from each CMA (in the first year, accounting for movers) with at least 10 total years of all the mentioned data, this should be sufficient to account for variation within subsets of socio-economic characteristics and mode choices. However, It is possible to go on about what a perfect data set could be for this analysis, increasing granularity of the panel data time frames could increase the accuracy of the results, a more complicated matter would be increasing the granularity of geographic areas from CMAs to census subdivisions CSDs and having data on all the associated characteristics of each CSD.

As mentioned in the literature review, many of the discrete choice modelling frameworks have a similar origin that is Johnson (1966), although he did have contemporaries who authored exceptional work on DCM models in transportation Johnson (1966) exemplifies the early thought particularly with respect to his analysis of time. Jara-Diaz (2000), and Correia et al. (2019) will provide the remainder of the framework that this paper will employ to develop a DCM approach to rail travel demand in the context of the Windsor – Quebec City corridor. The work of Correia et al. (2019) will prove particularly useful in our consideration of the potential shift in consumer preferences for rail vs. private vehicle travel as autonomous vehicle technology becomes more widespread. In this regard, bus, air, and rail transportation will have relatively far less innovation in the near future and therefore the value of travel time (VOTT) will also be unchanged, indicating no change in preference for them, ceteris paribus, the socioeconomic – demographic, CMA level TFF infrastructure factors.

The mathematics behind the DCM methodology is akin to utility maximization frameworks from microeconomics. From recent literature using Correia et al. (2019) and Jara-Diaz (2000) as examples, the mainstream method of estimating the demand for a travel mode with a DCM is to first develop a value of travel time (VOTT) measure for each of the competing modes of travel. This involves the utility maximization subject to time and income budget constraints as given by Jara-Diaz (2000) and Correia et al. (2019).

$$(5) \quad MaxU_k(G, L, W, \mathrm{J}_i)$$

Subject to,

$$(6) \quad G + c_i = wW$$

$$(7) \quad L + W + \mathrm{J}_i = \tau$$

$$(8) \quad \mathrm{L} \geq \alpha \mathrm{G}$$

Where U_k would be the utility of individual k, $(\mathrm{k} = 1, 2, \dots, \mathrm{n})$ for n individuals in the longitudinal survey cohort, G is consumption in monetary value, L is leisure expressed in time units, W is work expressed in time units, and J_i is the travel (journey) time between CMAs for each travel mode i, $\mathrm{i} = (\mathrm{bus}, \mathrm{rail}, \mathrm{private\ vehicle}, \mathrm{air})$. w is the salary per time unit, τ is the total available time, α is the consumption time per monetary unit of G. In the context of the panel data described earlier for the Windsor – Quebec City corridor, each term (5) – (8) would be evaluated for all individuals $(\mathrm{k} = 1, 2, \dots, \mathrm{n})$, as well each time period t in the panel data. In practice, (5) is a utility function of all factors effecting the arguments G, L, W, J_i, (6) and (7) are considered the income and time constraints respectively, and (8) is an expression of the fact that time to consume αG must occur during leisure time.

As shown in Correia et al. (2019) and Jara-Diaz (2000) we can now express a conditional maximization problem by replacing the equality constraints:

$$(9) \quad Max_W U_k[(wW - c_i), (\tau - W - \mathrm{J}_i), \mathrm{W}, \mathrm{J}_i]$$

Subject to,

$$(10) \quad \tau - W - \mathrm{J}_i - \alpha(wW - c_i) \geq 0$$

From this model we can get the VOTT for each individual k, a travel mode i, and for each time period t included assuming preferences do change over time.

$$(11) \qquad VOTT_{k,\mathrm{i,CMA}} = \frac{\frac{\partial V_i}{\partial t_i}}{\frac{\partial V_i}{\partial c_i}} = \mathrm{w} + \frac{\frac{\partial U_k}{\partial \mathrm{W}}}{\frac{\partial U_k}{\partial \mathrm{G}} - \theta\alpha} - \frac{\frac{\partial U_k}{\partial J_i}}{\frac{\partial U_k}{\partial \mathrm{G}} - \theta\alpha}$$

From (11), the term $\frac{\frac{\partial U_k}{\partial \mathrm{W}}}{\frac{\partial U_k}{\partial \mathrm{G}} - \theta\alpha}$ is the value of work for individual k in direct utility expressed in monetary units. $\frac{\frac{\partial U_k}{\partial J_i}}{\frac{\partial U_k}{\partial \mathrm{G}} - \theta\alpha}$ is the utility of traveling by itself expressed in monetary units (value of travelling as a commodity) with θ being the multiplier of constraint (8). $\mathrm{w} + \frac{\frac{\partial U_k}{\partial \mathrm{W}}}{\frac{\partial U_k}{\partial \mathrm{G}} - \theta\alpha}$ would be the of leisure, which would equal income for individual k in panel time period t, plus the utility of time at work transformed into monetary units. Therefore, the subjective value of travel time is the difference between the leisure value of time and the value of time as a commodity (Jara-Diaz, 2000; Correia et al., 2019).

Using these derivations from the past literature it is possible to develop a measure of willingness-to-pay for each individual and track the change in WTP over all our panel time periods. Again, subject to the ideal data conditions on socio-economic – demographic characteristics of all individuals as well as some additional CMA infrastructure characteristics (McFadden, 1974; Bradley, 1988; Orfeuil & Soleyret, 2002; Limtanakool et al., 2006a, 2006b), we can begin to formulate our regression specification to find the coefficients of all variables that affect the travel mode decision (and non-travel decision) of the average individual in our sample for each CMA in the corridor. The regression equation could then be specified for individual k, during time period t, and modal choice i, as follows:

$$(12) \qquad U_{k,\mathrm{t,i}} = \beta_x X_{\mathrm{t,i}} + \beta_\tau \tau_{k,\mathrm{t}} + \beta_\rho \rho_{k,\mathrm{t,CMA}} + \gamma_{k,\mathrm{t,i}}$$

Where β_x is a vector containing the estimated taste parameters associated with individual k, during time period t, for the subjectively desirable attributes of travel mode choice i, $X_{\mathrm{t,i}}$ is a vector of the attributes of that mode i, during panel time period t. β_τ is a vector that is the estimable parameter that captures the utility explained by socio-economic – demographic characteristics of individual k, during time period t, and $\tau_{k,\mathrm{t}}$ is vector containing the socio-economic – demographic characteristics of individual k, during time period t (Correia et al., 2019). β_{ρ} is a vector containing the estimable parameter that is the utility explained by CMA level infrastructure, and $\rho_{k,\mathrm{t,CMA}}$ is the CMA specific data on attributes such as quality of public transport (last-mile infrastructure), density of the CMA, etc. (McFadden, 1974; Bradley, 1988; Orfeuil & Soleyret 2002). Lastly, $\gamma_{k,\mathrm{t,i}}$ is a normalized vector for each individual k, where $\gamma_{k,\mathrm{t,i}} \epsilon [0{,}100]$ indicating the percentage of their travel during panel time period t, and for travel mode i, that is for the purpose of business, a value of 100 indicating that the individual only travelled inter-city for business purposes during the panel time period for a given mode (Limtanakool, 2006a, 2006b).

With this DCM framework, it would then be possible to run a regression specified for travel preferences for each mode, on average across time periods, for each of the 18 CMAs. Specifically, the prevailing choice by DCM modellers has been regressing a logit model that estimates the maximum likelihood of seeing the modal utility coefficients given the probability of having the sample data described, we could estimate the preference for each mode for each CMA. Once the results are deemed robust, it would be a simple matter of taking the weighted average preference for each mode in each CMA, to develop a set of 4 scalars that are the estimated percentage of each CMA population that would be the total demand for each of the 4 modes (bus, rail, private vehicle, air) during a given panel time period. These percentages could then be used in tandem with the gravity model results that already incorporated demographic projections into the expected demand for travel between each CMA pair.

Conclusion

An analysis of the literature on Windsor-Quebec corridor rail travel, begs the question of whether high-speed rail implementation is feasible, this study, if carried out quantitatively, can provide valuable insight on the analysis of potential demand given an introduction of high-speed rail service, using a combination of ad-hoc and discrete choice methodologies. Policymakers and private investors interested in the problem could modify the DCM taste preferences to account for increased speeds in rail service and the associated utility changes, they could then follow the effect on demand stemming from a shift toward rail service ceteris paribus. Other applications include the effect of increased autonomous vehicle accessibility, or a reduction in travel for the purpose of business in a post-COVID-19 era.

Importantly, the policymaker or investor could rely on these models as a tractable method of analysing the dynamic nature of inter-city travel demand and the causal factors that affect the utility derived from rail travel.

References

Anderson, J.E. (2011) The Gravity Model. *Annual Review of Economics*. 3, 133-60. https://doi.org/10.1146/annurev-economics-111809-125114

Becker, G.S. (1965). A Theory of the Allocation of Time. *The Economic Journal*. 75(299), 493-517. https://doi.org/10.2307/2228949

Ben-Akiva, M., Lerman, S.R. (1985) Discrete choice analysis: theory and application to travel demand. The MIT Press, Cambridge. https://mitpress.mit.edu/books/discrete-choice-analysis

Bradley, **M.** (1988) Realism and adaptation in designing hypothetical travel choice concepts. *Journal of Transport Economic Policy*. XXII:121–137. https://www.jstor.org/stable/20052838

Correia, G. H. de A., Looff, E., van Cranenburgh, S., Snelder, M., & van Arem, B. (2019). On the Impact of Vehicle Automation on the Value of Travel Time While Performing Work and Leisure Activities in a Car: Theoretical Insights and Results from a Stated Preference Survey. *Transportation Research: Part A: Policy and Practice*, 119, 359–382. https://doi.org/10.1016/j.tra.2018.11.016

Dargay, J.M., Clark, S. (2012) The determinants of long distance travel in great Britain. *Transportation Research A – Policy and Practice*. 46(3):576–587. https://doi.org/10.1016/j.tra.2011.11.016

Georggi, N.L., Pendyala, R.M. (2001). Analysis of long-distance travel behavior of the elderly and low income. *In: TRB Conference on Personal Travel: The Long and Short of It*. Washington, DC. https://trid.trb.org/view/686685

Jara-Díaz, S.R. (2000) Allocation and valuation of travel time savings. In: D. Hensher and K. Button (eds.), *Handbooks in Transport. Vol. 1: transport modelling*. Pergamon Press, Oxford, pp 303–318. https://repositorio.uchile.cl/handle/2250/172998

Johnson, M. B. (1966). Travel Time and the Price of Leisure. *Western Economic Journal*, 4(2), 135. https://login.proxy.hil.unb.ca/login?url=https://www.proquest.com/scholarly-journals/travel-time-price-leisure/docview/1297357395/se-2?accountid=14611

Jones, I.S. and Nichols, A.J. (1983) The Demand for Inter-City Rail Travel in the United Kingdom: Some Evidence. *Journal of Transport Economics and Policy*, 17(2), 133-153. https://www.jstor.org/stable/20052678

Limtanakool, N., Dijst, M., Schwanen, T. (2006a). On the participation in medium and long-distance travel: a decomposition analysis for the UK and The Netherlands. *Tijdschrift voor*

Economische en Sociale Geografie. 97(4), 389–404. https://doi.org/10.1111/j.1467-9663.2006.00347.x

Limtanakool, N., Dijst, M., Schwanen, T. (2006b). The influence of socioeconomic characteristics, land use and travel time considerations on the mode choice for medium- and longer-distance trips. *Journal of Transport Geography*. 14, 327–341. https://doi.org/10.1016/j.jtrangeo.2005.06.004

Mallett, W.J. (2001). Long-distance travel by low-income households. *In: TRB Conference on Personal Travel: The Long and Short of It*. Washington, DC. https://trid.trb.org/view/686689

McFadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics,* 3 303-328. https://eml.berkeley.edu/reprints/mcfadden/measurement.pdf

Mokhtarian, P.L., Salomon, I., Redmond, L.S. (2001). Understanding the demand for travel: it's not purely 'derived'. *Innovation: The European Journal of Social Science Research*. 14(4), 355–380. https://escholarship.org/uc/item/9v97r6zc

Nelligan, T. (1982). Via Rail Canada: The first five years. *PJT Publishing*. ISBN 0-937658-08-1.

Orfeuil, J., & Soleyret, D. (2002). What are the interactions between the short and long-distance travel markets? *Recherche Transports Sécurité.* 76, 208-221. https://doi.org/10.1016/S0761-8980(02)00013-4

Oum, T.H. and Gillen, D.W. (1983). The structure of intercity travel demands in Canada: Theory tests and empirical results. *Transportation Research Part B: Methodological*. 17(3), 175-191. https://doi.org/10.1016/0191-2615(83)90013-9

Owen, A.D. and Phillips G.D.A. (1987) The Characteristics of Railway Passenger Demand. An Econometric Investigation. *Journal of Transport Economics and Policy.* 21(3), 231-253. https://www.jstor.org/stable/20052820

Statistics Canada. (2019). Projected population, by projection scenario, age and sex, as of July 1 (x 1,000). Retrieved from *Statistics Canada.* https://doi.org/10.25318/1710005701-eng

Statistics Canada. (2006). 2006 Census analysis series. Retrieved from *Statistics Canada.* https://www12.statcan.gc.ca/census-recensement/2006/as-sa/index-eng.cfm

SNC-Lavlin et al. (1995). Quebec-Ontario High Speed Rail Project: Preliminary Routing Assessment and Costing Study, Final Report. Retrieved from: *Ministère des Transports - Gouvernement du Québec.* http://www.bv.transports.gouv.qc.ca/mono/0985915.pdf

Transport Canada, Ministère des Transports - Gouvernement du Québec, Ministry of Transportation of Ontario. (1995). Quebec-Ontario High Speed Rail Project, Final Report. Retrieved from: *Ministère des Transports - Gouvernement du Québec.* http://www.bv.transports.gouv.qc.ca/mono/0985548.pdf

Wardman, M., Lythgoe, W., and Whelan, G. (2007). Rail Passenger Demand Forecasting: Cross-Sectional Models Revisited. *Research in Transportation Economics*. 20(1), 119-152. https://doi.org/10.1016/S0739-8859(07)20005-8

Wong, B., Nurul Habib, K.M. (2015) Effects of accessibility to the transit stations on intercity travel mode choices in contexts of high speed rail in the Windsor–Quebec corridor in Canada. *Canadian Journal of Civil Engineering.* 42(11): 930-939. https://doi.org/10.1139/cjce-2014-0493