Immigrant Mode Choice in Canadian Cities: Mixed Logit with GPS Panel Data and Joint RP-SP Estimation

Tareq Alsaleh*[1], Bilal Farooq[1], and Zachary Patterson[2]

[1]Laboratory of Innovations in Transportation (LiTrans), Toronto Metropolitan University, Canada
[2]Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Canada

Short summary

This study investigates how immigrant status and integration level shape transportation mode choice in Canadian cities. The analytical framework combines four elements: a continuous multidimensional integration index, mixed logit models with systematic immigrant-linked taste heterogeneity, joint revealed–stated preference estimation, and a GPS-based custom mobile app panel dataset of approximately 80,000 trip observations and 622 stated preference scenarios from 100 participants in Toronto and Montréal. A multi-stage data processing pipeline transforms raw GPS traces into estimation-ready choice sets with decomposed transit journey attributes. The mixed logit specification reveals that sampled immigrants exhibit approximately 66% lower sensitivity to in-vehicle travel time than Canadian-born residents, while higher integration is associated with reduced transit use. Counterfactual simulations demonstrate that access time reductions are approximately three times more effective than fare reductions in shifting car commuters to transit. Five-fold cross-validation yields 80–82% prediction accuracy across specifications.
Keywords: Discrete choice modelling, immigrant travel behaviour, mixed logit, GPS panel data, revealed and stated preference, integration index

1 Introduction

Canada is one of the world's foremost immigrant-receiving nations, with immigrants comprising 23% of the total population as of the 2021 Census, the highest proportion among G7 nations Immigration, Refugees and Citizenship Canada (2022); Statistics Canada (2022). The metropolitan areas of Toronto and Montr'eal receive a disproportionate share of annual admissions, placing direct pressure on urban transportation systems. Yet the relationship between immigrant status and mobility is bidirectional, while transportation access enables integration into employment, healthcare, education, and social life Delbosc & Shafi (2023), the process of integration itself reshapes how immigrants travel. As immigrants acquire language proficiency, obtain driver's licences, and expand social networks, their mode choices evolve in ways that cross-sectional analyses cannot capture Welsch et al. (2018); Asgari et al. (2017).

A growing body of evidence demonstrates that immigrants do, in fact, travel differently from the native-born population. Studies across North America and Europe consistently find higher reliance on public transit and lower drive-alone rates, particularly during early settlement, with these differences diminishing through "transport assimilation" Chatman & Klein (2009); Asgari et al. (2017). Smart (2015) demonstrated that immigrant concentration at the neighbourhood level is associated with higher transit and active mode shares, an effect extending to non-immigrant residents. Xu (2018) confirmed the assimilation pattern using repeated cross-sectional U.S. data, while Harun et al. (2021) documented a positive association between immigrant concentration and transit commute shares in Toronto. Welsch et al. (2018) found that duration of stay was strongly associated with car use among immigrants in Germany, and Welsch & Mattioli (2025) showed that persistent ethnic differences in bus use remained even after adjusting for socioeconomic mediators in England.

Despite this evidence, three methodological limitations persist. First, immigrant status is overwhelmingly treated as a binary variable, obscuring the multidimensional nature of integration Delbosc & Shafi (2023). Second, most studies rely on cross-sectional surveys that cannot capture within-individual variation. Third, the multinomial logit remains the dominant specification Asgari et al. (2017); Chen et al. (2015), imposing taste homogeneity. Mixed logit relaxes these restrictions

Train (2009), yet no study has specified random parameters with systematic immigrant-driven taste heterogeneity. Similarly, joint RP-SP estimation Ben-Akiva & Morikawa (1990); Brownstone et al. (2000) has not been applied to immigrant populations with integration as a moderating variable.

This paper addresses these gaps through four contributions: (1) a continuous integration index weighted across economic (40%), social (30%), civic (20%), and health (10%) dimensions; (2) mixed logit models with systematic immigrant-driven heterogeneity in time sensitivity; (3) joint RP-SP estimation and explicit scale parameter analysis; and (4) a GPS smartphone panel collecting approximately 80,000 trip observations and 622 SP scenarios from 100 individuals over multiple months.

2 Methodology

Data collection and study design

Data collection was conducted between October 2024 and November 2025 in the Greater Toronto and Greater Montréal Areas using a custom-built mobile application, *BDMobility* Alsaleh et al. (2025); Otchere et al. (2025). These two regions were selected because they collectively receive around 40% of Canada's immigrant admissions and exhibit substantial variation in transit infrastructure, urban form, and linguistic context. The app serves three integrated functions (Figure 1): semi-passive GPS-based travel diary recording via the MotionTag platform, administration of static survey instruments, and dynamic generation of stated preference scenarios pivoted on participants' own observed trips. This approach balances the data richness of passive GPS tracking with the validation benefits of respondent engagement, and addresses the accuracy limitations identified in purely automated systems (Harding et al., 2021; Marra et al., 2019; Azoulay & Patterson, 2024).

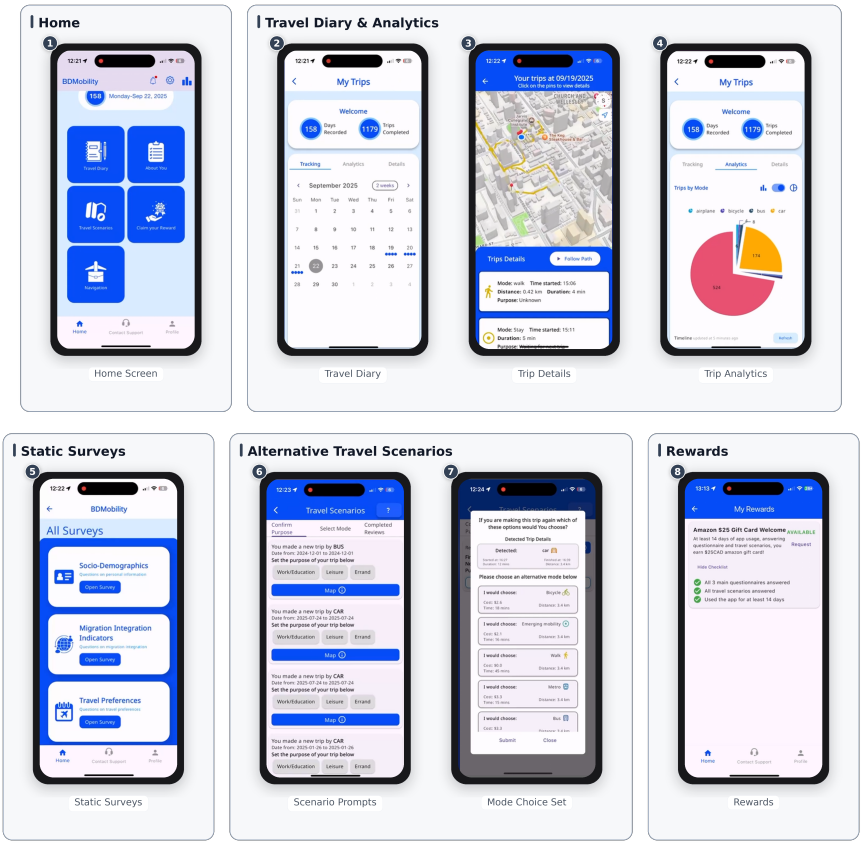


Figure 1: BDMobility application user interface: travel diary, survey instruments, and dynamic SP scenario generation.

A deliberate methodological choice is the prioritization of longitudinal depth over cross-sectional breadth. The multi-month panel generates a median of approximately 107 observed trips per

participant, capturing day-to-day mode switching, weather responses, and routine versus non-routine travel that single-day surveys cannot detect. The trade-off is a smaller number of unique individuals, but the objective is to identify mechanisms rather than to calibrate a regional demand model.

The integration index operationalizes the Canadian Index for Measuring Integration (CIMI) framework Canadian Index for Measuring Integration (CIMI) (2020) at the individual level through a 37-question survey capturing economic integration (employment, wage alignment), social integration (community ties, belonging), civic integration (political participation, volunteering), and health integration (healthcare access, life satisfaction). Crucially, the index is applied to first-generation immigrants (born outside Canada), second-generation (born in Canada with at least one foreign-born parent), third-generation or higher (both parents also born in Canada), capturing a continuous embeddedness gradient that varies meaningfully within both populations.

The SP component employs a decision-gate algorithm that selects eligible trips based on distance, observed mode, frequency, and purpose, then dynamically generates alternative-mode attributes pivoted on actual trip characteristics and local transit provider data Alsaleh et al. (2025). Unlike conventional standalone SP surveys, these in-app scenarios are grounded in real travel experience, reducing the cognitive burden of evaluating unfamiliar attribute levels. This design yielded 622 SP observations across the sample.

Data processing pipeline

The raw data produced by the mobile app required extensive post-processing. A multi-stage pipeline transforms raw GPS traces, travel diary records, and survey responses into estimation-ready choice sets. Raw trajectories were validated against study area boundaries. A notable challenge arose from the monetary incentive structure, which attracted fraudulent enrolments from outside Canada. A multi-layered screening protocol addressed this through IP geolocation cross-referencing, trajectory diagnostics flagging impossible speeds and zero-variance coordinates, and internal consistency checks on survey responses. All 100 retained participants passed every stage, which highlights the growing need for robust validation protocols in incentivized app-based surveys.

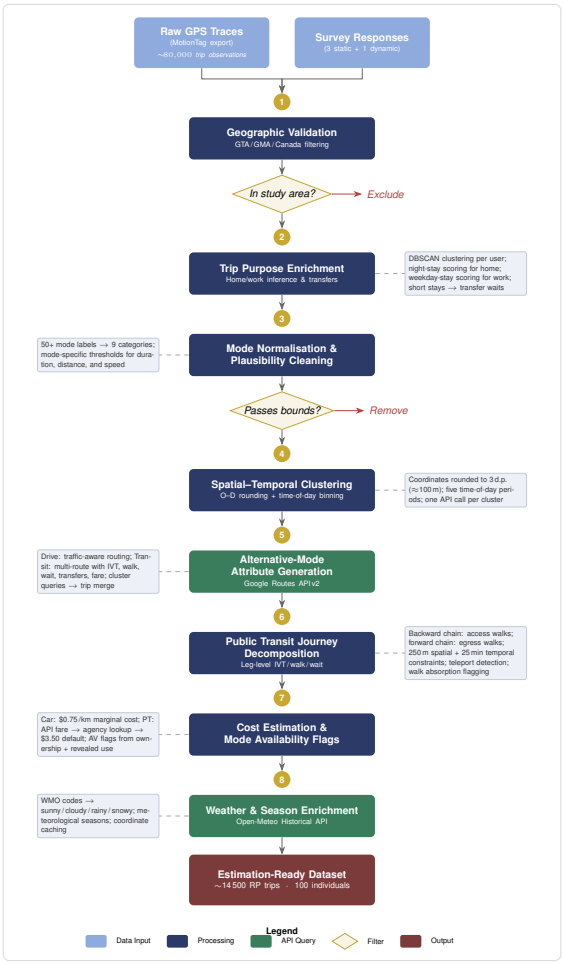
Trip purpose was enriched through a density-based spatial clustering algorithm that identified home and work anchor locations per participant using multi-criteria scoring. For the choice set reformulation of the observed trips, the alternative-mode attributes were generated by rounding origins and destinations to three decimal places (100 m) and binning departure times into five periods, then querying the Google Routes API for each unique spatiotemporal cluster under traffic-aware conditions. For observed transit trips, the raw GPS event stream was analyzed at the individual leg level to decompose journeys into access walk, platform wait, in-vehicle time, and egress walk, with spatial continuity and temporal plausibility constraints enforced at each transition. Car operating costs were calculated at a marginal rate of \$0.75 CAD/km, and transit fares were derived from a three-tier hierarchy: API-returned fares, a curated 60-agency lookup table, or a \$3.50 CAD default. Weather codes from the Open-Meteo API and meteorological season were appended to each record. Figure 2 presents the overall pipeline structure and a worked example illustrating how a single observed bus trip is decomposed into an estimation-ready choice set.

Discrete choice modelling framework

The analysis employs a 2×2 model comparison framework crossing model structure (MNL vs. MXL) with data source (RP-only vs. joint RP-SP), yielding four specifications that isolate the contribution of each methodological dimension. The MNL baseline specifies systematic utility for seven alternatives (car, bus, subway, train, walk, bicycle, and e-mobility in joint models), with car and train as reference alternatives. Table 1 defines the variables entering the utility specifications. The core utility functions are:

$$
\begin{aligned}
V_{\text{car}} &= \beta_C \,\text{Cost}_s + \beta_T \,\text{IVTT}_s + \beta_{D1} \,\text{Dist}_s + \beta_W \,\text{TP_WS} \\
V_{\text{bus}} &= \alpha_{\text{bus}} + \beta_C \,\text{Cost}_s + \beta_T \,\text{IVTT}_s + \beta_A \,\text{Walk}_s + \beta_{D2} \,\text{Dist}_s + \beta_{Ch} \,\text{CHILD} + \beta_{I1} \,\text{INTEG_C} \\
V_{\text{sub}} &= \alpha_{\text{sub}} + \beta_C \,\text{Cost}_s + \beta_T \,\text{IVTT}_s + \beta_A \,\text{Walk}_s + \beta_F \,\text{FT} + \beta_{Ch} \,\text{CHILD} + \beta_M \,\text{MIG} + \beta_{I1} \,\text{INTEG_C} \\
V_{\text{walk}} &= \alpha_{\text{walk}} + \beta_T \,\text{TT}_s + \beta_{Sw} \,\text{STU} + \beta_{D4} \,\text{Dist}_s + \beta_{I2} \,\text{INTEG_C} + \beta_{Sn} \,\text{SNOW}
\end{aligned} \tag{1}
$$

The mixed logit relaxes taste homogeneity through two random parameters. In-vehicle travel time



(a) Data processing pipeline from raw GPS traces to estimation-ready dataset. Diamond nodes indicate quality filters.



(b) Worked example: transit journey decomposition for a single observed bus trip (Toronto, AM peak). Left: GPS event stream with decomposition; Right: estimation-ready choice set.

Figure 2: Data processing pipeline and worked example. a Overview of the multi-stage processing workflow; b a single bus trip traced through the pipeline, showing how raw GPS events are decomposed into access walk, platform wait, in-vehicle time, and egress walk, and how unchosen alternative attributes are generated via the Google Routes API.

Table 1: Variable definitions

Variable	Description	Enters utility of
Level-of-service attributes (continuous, scaled /10)		
Cost_s	Trip cost (CAD/10)	All motorized modes
IVTT_s	In-vehicle travel time (min/10)	All modes
Walk_s	Walk/access time (min/10)	Bus, subway, train
Trip characteristic (continuous, scaled /1,000)		
Dist_s	Trip distance (km/1000)	Mode-specific
Sociodemographic characteristics (binary)		
TP_WS	Work or study trip (1 = yes)	Car
CHILD	Children aged 0–10 (1 = yes)	Bus, subway
FT	Full-time employed (1 = yes)	Subway
STU	Student (1 = yes)	Train, walk
MIG	Foreign-born (1 = yes)	Subway (MNL only)
SNOW	Snow conditions (1 = yes)	Walk, bicycle
Integration measure (continuous, mean-centred)		
INTEG_C	Integration index (1–10, centred at $\bar{x}$)	PT modes; active (MNL)

follows a normal distribution with systematic heterogeneity linked to immigrant status:

$$\beta_{T,n} = \big(\mu_T + \delta_{\text{MIG}} \cdot \text{MIG}_n\big) + \sigma_T \cdot z_n, \quad z_n \sim \mathcal{N}(0,1) \tag{2}$$

where μ_T is the population mean, δ_{MIG} captures the systematic shift for immigrants, and σ_T captures residual unobserved heterogeneity. Cost follows a negative lognormal distribution to enforce the theoretically correct sign Hensher & Greene (2003):

$$\beta_{C,n} = -\exp\big(\mu_C + \sigma_C \cdot z_n\big), \quad z_n \sim \mathcal{N}(0,1) \tag{3}$$

The MXL choice probability integrates over the distribution of random parameters, exploiting the panel structure through the product over each individual's choice occasions, approximated using 500 Halton draws. Joint RP-SP models share taste parameters with a scale parameter μ_{SP} accommodating differential error variances: $V_{\text{SP}} = \mu_{\text{SP}} \cdot V(\boldsymbol{\beta})$ Ben-Akiva & Morikawa (1990). All models are estimated using the Apollo package Hess & Palma (2019) in R, with five-fold cross-validation for out-of-sample assessment.

3 Results and discussion

Sample and travel diary

The final sample comprises 100 individuals (33% first-generation immigrants, 27% second-generation, 41% third + generation). where, first-generation immigrants are born outside Canada, second-generation are born in Canada with at least one foreign-born parent and third-generation or higher are born in Canada with both parents also born in Canada. After the processing pipeline, approximately 14,500 RP trips and 622 SP scenarios were retained. Figure 3 presents the spatial distribution of recorded trip origins, and Figure 4 summarizes the sociodemographic profile by immigrant status. Canadian-born (second and third + generations) participants cluster in the 45–54 age bracket (49%), whereas immigrants (first generation) are more evenly distributed across working ages. Full-time employment is higher among Canadian-born (72% vs. 38%), while 47% of immigrants are students. Car ownership is high overall (83%), though immigrants are substantially more likely to be car-free (41% vs. 6%).

Table 2 summarizes the trip diary. Walking dominates trip count (49.7%), a distinctive feature of multi-day GPS panels where continuous recording captures the routine short trips that participants omit from single-day diaries. When measured by distance, car dominates (66.1%). Canadian-born participants exhibit higher car mode share (30% vs. 26%), while immigrants show higher walking (53% vs. 47%) and cycling (4% vs. 1%).

The integration index reveals a clear gradient between groups (Figure 5): Canadian-born participants report a mean of 8.5 versus 6.7 for immigrants on the 1–10 scale, with the largest gap in civic integration (6.0 vs. 8.4) and the smallest in health (6.8 vs. 8.5). The SP scenarios were triggered

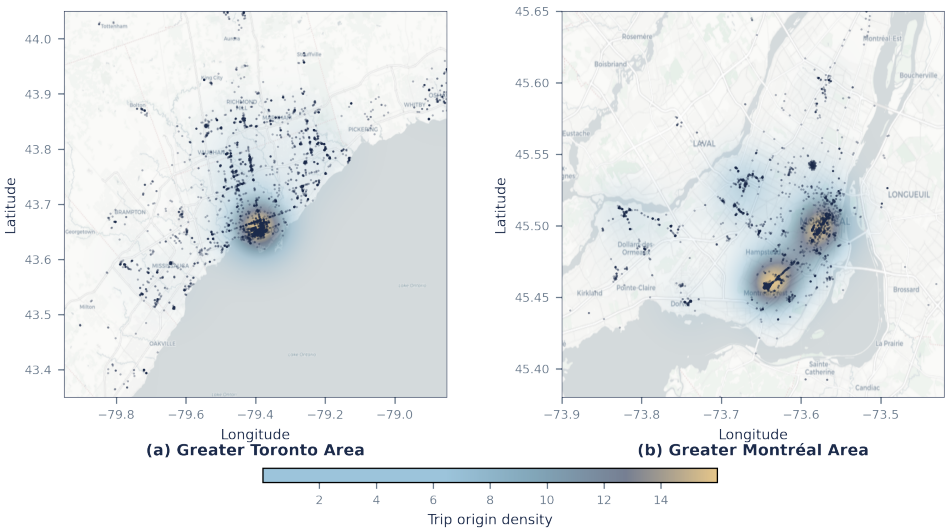
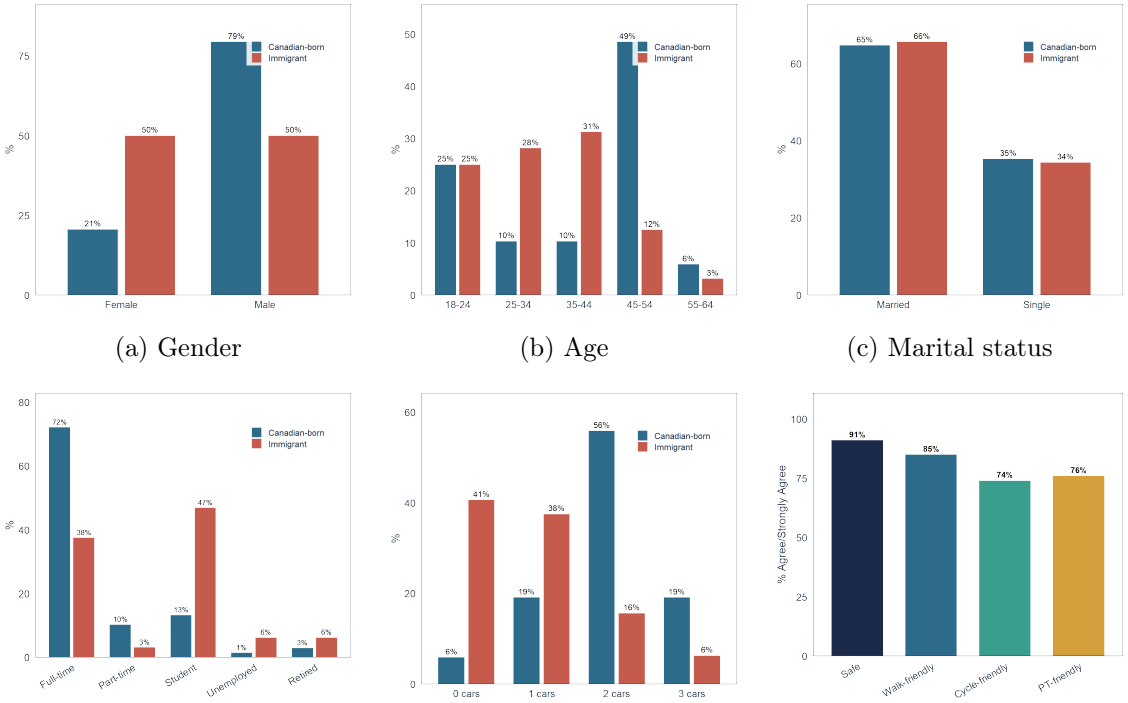


Figure 3: Spatial distribution of recorded trip origins: (a) Greater Toronto Area and (b) Greater Montréal Area. Colour intensity reflects trip origin density.



(a) Gender (b) Age (c) Marital status

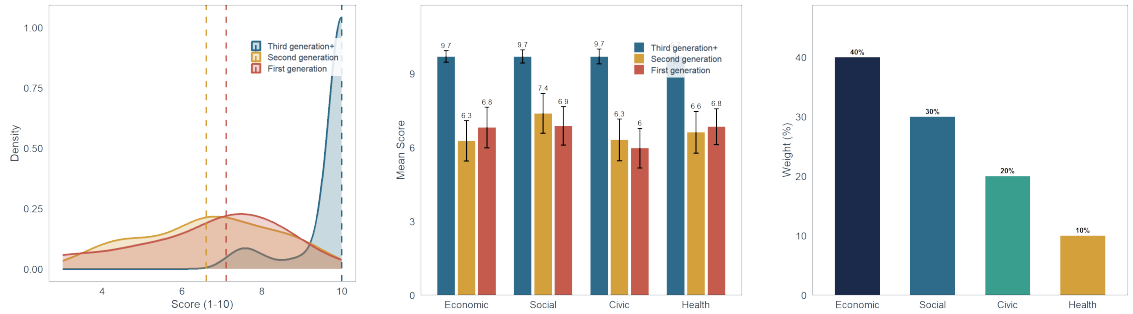
(d) Employment (e) Car ownership (f) Residential perceptions

Figure 4: Sociodemographic profile of study participants by immigrant status.

Table 2: Trip attributes by mode

Mode	Trips	Share (%)	Dist. mean (km)	Time mean (min)	Cost (CAD)
Car	4,979	28.0	13.5	20.7	13.90
Bus	1,461	8.2	4.2	13.9	3.50
Subway	1,713	9.6	5.2	12.6	3.50
Train	379	2.1	32.6	38.2	3.50
Walk	8,841	49.7	0.6	9.3	0.00
Bicycle	411	2.3	3.0	12.9	0.00

dynamically based on participants' observed trips, with half generated from car trips exceeding 5 km and the remainder from shorter car and public transit trips. Overall, 52.6% of respondents chose to retain their current mode. Disaggregation by original observed mode and immigrant status reveals distinct switching patterns. When the original trip was by car, Canadian-born participants overwhelmingly retained car (81%), while immigrants showed modestly greater openness to alternatives. When the original trip was by bus, Canadian-born participants were more likely to retain bus (65% vs. 54% for immigrants). The starkest contrast emerges for subway trips: immigrants strongly retained subway (73%) whereas Canadian-born participants showed a broader distribution across alternatives (54% subway retention, 35% shifting to walk).



(a) Integration density (b) Sub-dimensions (c) Dimension weights

Figure 5: Integration index: a density by group with median lines, b sub-dimension means with 95% CIs, and c dimension weights.

Model estimation and validation

Table 3 summarizes the estimation results. All models exhibit substantial improvement over the null model, with adjusted ρ^2 values ranging from 0.455 to 0.532. The MXL models achieve higher adjusted ρ^2 than their MNL counterparts despite having fewer parameters, which indicates that the inclusion of random parameters captures meaningful unobserved preference heterogeneity. Five-fold cross-validation yields mean prediction accuracy of 80–82% with standard deviations below 1%, and the key behavioural parameters are stable across all folds.

Table 4 reports the MXL parameter estimates. All four specifications yield behaviourally consistent responses with respect to cost, travel time, and distance. Walk and access time coefficients are consistently larger in magnitude than in-vehicle time by approximately 28–31%, which highlights a higher behavioural penalty for out-of-vehicle components, consistent with the established transport economics literature Wardman (2004). The distance coefficients reveal a clear modal hierarchy: longer trips increase the relative attractiveness of car and train, while substantially reducing the likelihood of active modes.

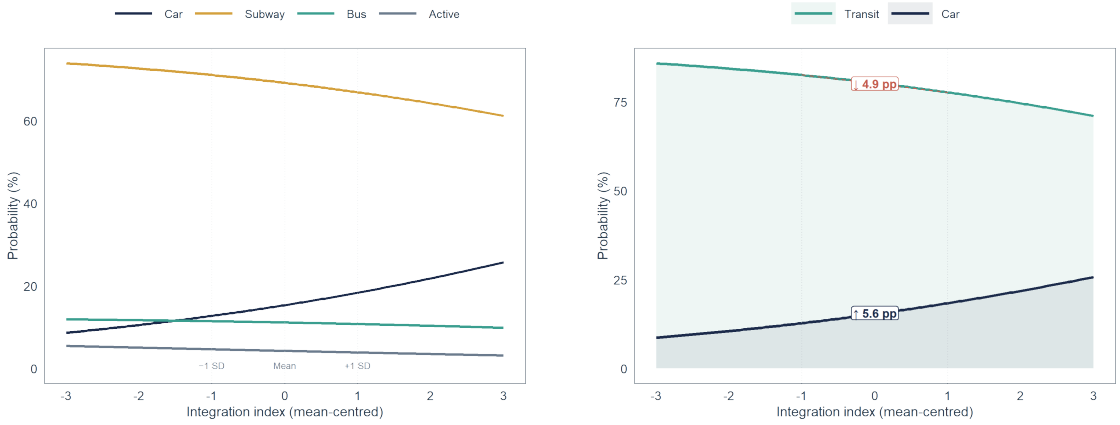
In both MXL specifications, the immigrant shift parameter (δ_{MIG}) is positive and statistically significant, with estimates of 0.556 in M2 (robust $t = 5.25$) and 0.525 in M4 (robust $t = 3.58$). This shift reduces the magnitude of the IVTT coefficient for immigrants, resulting in mean values of -0.292 (M2) and -0.273 (M4), compared to -0.848 and -0.798 for Canadian-born individuals. This corresponds to an approximate 66% reduction in time sensitivity and an immigrant-to-Canadian-born value-of-travel-time ratio of 0.34 in both specifications. The estimated standard deviation ($\sigma_T \approx 0.71$) is of similar magnitude to the population mean, suggesting additional unobserved heterogeneity beyond the immigrant effect. The MNL specifications yield overall VOT

Table 3: Model fit comparison

	M1 MNL RP	**M2** MXL RP	**M3** MNL RP-SP	**M4** MXL RP-SP
Parameters	21	17	24	19
Observations	14,502	10,241	15,124	9,455
$\mathcal{LL}$(final)	−6,842.8	−4,257.6	−7,487.5	−4,254.9
Adj. $\rho^2(0)$	0.473	0.532	0.455	0.504
AIC	13,728	8,549	15,023	8,548
μ_{SP}	—	—	0.298	0.281
Five-fold observation-level cross-validation				
CV accuracy (mean ± SD)	82.0±0.7%	79.9±0.7%	81.8±0.8%	79.6±0.7%

MXL observations reflect stratified trip caps (300 per person) to prevent panel-product underflow.

estimates of \$26–28 CAD/hr for in-vehicle time and \$34–36 CAD/hr for walk/access time. The integration index (β_{I1}) is negative and highly significant in both MNL models ($t = -9.03$ in M1; $t = -10.18$ in M3), indicating that higher integration is associated with reduced public transit utility relative to car. In the MXL specifications, however, this coefficient loses significance ($t = -0.55$ in M2; $t = -0.73$ in M4) while retaining its negative sign. This pattern is expected rather than contradictory: the random cost and time parameters in the MXL capture individual-level distributional heterogeneity that the fixed-coefficient MNL can only approximate through explicit interaction terms such as the integration index. The fact that the integration effect is absorbed by the random parameters is itself informative, as it suggests that integration partly operates through the same preference dimensions, time and cost sensitivity, that the MXL already models as continuously varying across individuals. A natural concern is whether the integration gradient is simply proxying for income, since more integrated immigrants tend to have higher earnings. A sensitivity analysis addresses this directly: introducing income into the M1 specification alongside the integration index leaves β_{I1} unchanged (-0.179, $t = -9.02$ with income vs. -0.179, $t = -9.03$ without), confirming that the index captures dimensions of socioeconomic embeddedness that are empirically distinct from income. Figure 6 illustrates how this gradient translates into mode choice probabilities under representative travel conditions.



(a) All mode probabilities

(b) Car vs transit

Figure 6: Integration effect on mode choice probabilities (Model 3, representative 10 km work commute).

The joint RP-SP models estimate a scale parameter μ_{SP} that captures the relative error variance between the two data sources. The estimated values (0.298 in M3; 0.281 in M4) are well below unity, implying that SP error variances are 3.4 to 3.6 times larger than those of RP. At face value, this would suggest substantial hypothetical bias in the stated preference responses. However, a distinctive feature of this dataset is the extreme observation imbalance between the two sources: approximately 14,500 RP observations versus only 622 SP, a ratio exceeding 23:1. In joint estimation, this imbalance means that the RP component overwhelmingly dominates the likelihood

Table 4: MXL parameter estimates: Model 2 (RP) and Model 4 (Joint RP-SP)

	M2: MXL RP		**M4: MXL RP-SP**	
Parameter	Est.	t	Est.	t
Scale parameter				
μ_{SP}	—	—	0.281	2.76^{***}
Alternative-specific constants				
α_{bus}	−0.396	−1.16	−0.519	−1.50
α_{sub}	2.107	6.07^{***}	2.016	5.66^{***}
α_{walk}	1.088	3.00^{***}	0.834	2.27^{**}
α_{bike}	−1.713	-3.88^{***}	−1.869	-4.77^{***}
α_{emob}	—	—	−0.810	−0.35
Random parameters				
μ_T (IVTT mean)	−0.848	-8.57^{***}	−0.798	-7.21^{***}
σ_T (IVTT s.d.)	0.713	11.25^{***}	0.706	9.81^{***}
δ_{MIG} (immigrant shift)	0.556	5.25^{***}	0.525	3.58^{***}
μ_C (Cost log-mean)	−2.130	-3.97^{***}	−2.039	-5.33^{***}
σ_C (Cost log-s.d.)	1.724	6.84^{***}	1.697	9.10^{***}
Fixed parameters				
β_A (Walk time)	−0.448	-2.87^{***}	−0.531	-4.55^{***}
β_{D1} (Dist. Car)	0.394	10.46^{***}	0.382	8.95^{***}
β_{D2} (Dist. PT)	0.110	1.86^{*}	0.107	1.80^{*}
β_{D3} (Dist. Train)	0.276	7.04^{***}	0.251	5.60^{***}
β_{D4} (Dist. Active)	−0.864	-4.64^{***}	−0.846	-5.30^{***}
β_F (FT → Sub)	−0.684	-2.10^{**}	−0.673	-2.06^{**}
β_{Sw} (STU → Walk)	0.559	2.10^{**}	0.656	2.20^{**}
β_{I1} (INTEG_C → PT)	−0.095	−0.55	−0.106	−0.73

500 Halton draws. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Robust t-statistics. Car and train ASCs normalized to 0.

function, and the scale parameter is forced to absorb the statistical weight differential between the two samples in addition to any genuine differences in response reliability. To disentangle sample imbalance from true hypothetical bias, the joint MNL was re-estimated on a balanced subsample comprising only the RP trips that had originally triggered SP scenarios, with sociodemographic parameters held fixed at their full-sample Model 3 estimates. Under this balanced specification, μ_{SP} increases substantially to 0.491 ($t = 4.69$), nearly doubling relative to the full-sample estimate. This indicates that approximately two-thirds of the apparent SP noise in the full-sample estimation is an artefact of the observation imbalance rather than a reflection of genuine hypothetical bias. The residual gap from unity is consistent with the moderate hypothetical bias typically observed in SP studies, and the in-app pivoted design, which grounds each scenario in the participant's own observed trip attributes, likely contributes to the relatively high SP reliability. This decomposition carries broader methodological implications for joint RP-SP estimation: in datasets where the RP sample is orders of magnitude larger than the SP sample, a low μ_{SP} should not be interpreted as evidence of poor SP quality without first accounting for the influence of sample imbalance on the estimated scale.

Policy simulations

Building on these behavioural patterns, counterfactual simulations translate the estimated parameters into policy-relevant outcomes. Table 5 examines two intervention levers for a representative immigrant making a 10 km work commute: fare reduction and access time reduction. Eliminating transit fares increases transit probability by 3–4 percentage points, whereas eliminating a 15-minute access time produces gains of 10–12 percentage points, approximately three times the fare effect. This asymmetry reflects the higher behavioural penalty for out-of-vehicle time: reductions in access time simultaneously improve the perceived utility of all transit alternatives. The access time lever is particularly effective for highly integrated immigrants near the car–transit margin, who exhibit the largest gains. Critically, the integration-driven gap in transit probability persists under all simulated scenarios, which suggests that infrastructure improvements alone cannot fully offset the progressive shift toward car dependence that accompanies integration.

Table 5: Transit probability (%) under counterfactual scenarios, Model 3

Scenario	**Low integ.** (−1 SD)	**Avg integ.** (Mean)	**High integ.** (+1 SD)
Fare reduction			
Current ($3.25)	82.6	80.4	77.7
Free ($0.00)	86.0	84.1	81.9
Gain (pp)	*+3.4*	*+3.7*	*+4.2*
Access time reduction (fare = $3.25)			
15 min (baseline)	77.7	75.1	72.0
0 min	87.4	85.6	83.6
Gain (pp)	*+9.7*	*+10.5*	*+11.6*

Representative immigrant, 10 km work commute, full-time employed.

4 Conclusions

This study developed analytical framework for analyzing immigrant mode choice that integrates four elements not previously combined in a single analysis: a continuous multidimensional integration index, mixed logit models with systematic immigrant-driven taste heterogeneity, joint RP-SP estimation, and GPS-based smartphone panel data. A multi-stage data processing pipeline, encompassing fraud detection, transit journey decomposition, and routing-API-based alternative attribute generation, transforms raw GPS traces into estimation-ready choice sets.

Three core paterns emerge from the estimated models. First, the sampled immigrants exhibit approximately 66% lower sensitivity to in-vehicle travel time than Canadian-born residents, with a consistent VOT ratio of 0.34 across both MXL specifications. Second, higher integration is associated with reduced transit use at a rate of approximately 5 percentage points per standard deviation, an effect that persists under all simulated service improvement scenarios. Third, access time reductions are approximately three times more effective than fare elimination in shifting car commuters to transit, with the effect most pronounced for highly integrated immigrants near the car–transit decision margin.

The 2×2 model comparison confirms that these results are not artefacts of particular modelling choices: the core findings hold across model structures and data sources, with five-fold cross-validation yielding 80–82% prediction accuracy. The persistence of the integration-driven car adoption gradient under all simulated scenarios, including complete fare elimination, indicates that service improvements alone cannot fully retain transit users. International migration represents one of the most significant disruptions to an individual's mobility biography, temporarily unfreezing habitual behaviour and creating receptivity to new travel options. The estimated models suggest this window may be quantifiable. Immigrants exhibit substantially lower time sensitivity, greater willingness to explore alternatives in stated preference scenarios, and a transit-oriented baseline that erodes as integration progresses. Interventions delivered during the early settlement period, when these characteristics are most pronounced, are therefore likely to yield larger and more lasting effects than the same interventions targeted at established residents with entrenched car habits. For transit agencies in immigrant-receiving cities, this points toward early-stage mobility packages, investment in first/last-mile access rather than fare subsidies, and integration-aware demand management that recognizes the settlement period as a critical window for sustainable travel behaviour formation.

Acknowledgements

This research is funded by the Canada First Research Excellence Fund under the Bridging Divides program and the Canada Research Chair in Disruptive Transportation Technologies and Services (CRC–2021–0048). The authors thank Bara Rababah, Henrique De Freitas Serra, Isaac Otchere, and Thomas Zhao for their support in data collection.

References

Alsaleh, T., Otchere, I., Farooq, B., Patterson, Z., & Cherchi, E. (2025). Do perceptions migrate

with us? an investigation using automated travel scenarios generation from semi-passive travel diary recording. *Transportation Research Procedia*. (In press)

Asgari, H., Zaman, N., & Jin, X. (2017). Understanding immigrants' mode choice behavior in Florida: Analysis of neighborhood effects and cultural assimilation. *Transportation Research Procedia*, *25*, 3083–3099. doi: 10.1016/j.trpro.2017.05.319

Azoulay, B., & Patterson, Z. (2024). Towards the standardization of reporting in smartphone travel surveys: the development and application of the smartphone survey reporting guidelines (ssrgs). *Transportation research procedia*, *76*, 574–585.

Ben-Akiva, M., & Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General*, *24*(6), 485–495. doi: 10.1016/0191-2607(90)90015-X

Brownstone, D., Bunch, D. S., & Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, *34*(5), 315–338. doi: 10.1016/S0191-2615(99)00031-4

Canadian Index for Measuring Integration (CIMI). (2020). *Cimi user guide* (Tech. Rep.). Canadian Index for Measuring Integration. Retrieved from `https://integrationindex.ca/wp-content/uploads/2020/04/CIMI_UserGuide_EN_Final.pdf`

Chatman, D. G., & Klein, N. (2009). Immigrants and travel demand in the united states: Implications for transportation policy and future research. *Public Works Management & Policy*, *13*(4), 312–327.

Chen, H., Gan, Z., & He, Y. (2015). Choice model and influencing factor analysis of travel mode for migrant workers: Case study in Xi'an, China. *Discrete Dynamics in Nature and Society*, *2015*, 236216. doi: 10.1155/2015/236216

Delbosc, A., & Shafi, R. (2023). What do we know about immigrants' travel behaviour? A systematic literature review and proposed conceptual framework. *Transport Reviews*, *43*(5), 914–934. doi: 10.1080/01441647.2023.2180867

Harding, C., Faghih Imani, A., Srikukenthiran, S., Miller, E. J., & Habib, K. N. (2021). Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys. *Transportation*, *48*, 2433–2460.

Harun, R., Filion, P., & Moos, M. (2021). The immigrant effect on commuting modal shares: Variation and consistency across metropolitan zones. *Journal of Urbanism*, *15*(4), 421–441. doi: 10.1080/17549175.2021.1893798

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, *30*(2), 133–176.

Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of choice modelling*, *32*, 100170.

Immigration, Refugees and Citizenship Canada. (2022). *2022 annual report to Parliament on immigration* (Tech. Rep.). Government of Canada.

Marra, A. D., Becker, H., Axhausen, K. W., & Corman, F. (2019). Developing a passive GPS tracking system to study long-term travel behavior. *Transportation Research Part C: Emerging Technologies*, *104*, 348–368. doi: 10.1016/j.trc.2019.05.006

Otchere, I., Alsaleh, T., De Freitas Serra, H., Patterson, Z., & Farooq, B. (2025). Smartphone travel survey-based adaptive mode choice experiment. *Transportation Research Procedia*. (In press)

Smart, M. J. (2015). A nationwide look at the immigrant neighborhood effect on travel mode choice. *Transportation*, *42*, 189–209.

Statistics Canada. (2022, October). *Immigration and ethnocultural diversity: Key results from the 2021 Census* (Tech. Rep.). Statistics Canada.

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511805271

Wardman, M. (2004). Public transport values of time. *Transport policy*, *11*(4), 363–377.

Welsch, J., Conrad, K., & Wittowsky, D. (2018). Exploring immigrants' travel behaviour: Empirical findings from Offenbach am Main, Germany. *Transportation*, *45*, 733–750. doi: 10.1007/s11116-016-9748-9

Welsch, J., & Mattioli, G. (2025). The impact of migration background and ethnicity on car, bus and bicycle use in England. *Transportation*. (Advance online publication) doi: 10.1007/s11116-025-10696-5

Xu, D. (2018). Transportation assimilation revisited: New evidence from repeated cross-sectional survey data. *PloS one*, *13*(4), e0194296.