A Dynamic Latent Class Model to Estimate the Crowding Cost of Subway Riders

5 February 2020

Prateek Bansal
prateek.bansal@imperial.ac.uk

Daniel Hörcher
d.horcher@imperial.ac.uk

Daniel J. Graham
d.j.graham@imperial.ac.uk

Transport Strategy Centre, Department of Civil and Environmental Engineering
Imperial College London, UK

1 Introduction

Quantitative measurement of crowding disutilities in public transport is important in investment appraisals, demand modeling, and supply-side decisions such as fare optimisation. To this end, most of the studies use a stated preference (SP) survey and estimate the traveller's perceived value of crowding in terms of a *crowding multiplier* – the ratio of value-of-travel-time under crowded and uncrowded conditions (Wardman and Whelan, 2011). SP studies generally elicit preferences of riders in a hypothetical route choice experiment and estimate discrete choice models (DCMs) to obtain the crowding multiplier (See Bansal et al., 2019, for the review). Whereas the hypothetical bias is a major limitation of the SP data, the required information to estimate DCMs (riders' route preferences and attributes of all available routes) is difficult to obtain using conventional revealed preference (RP) surveys (Tirachini et al., 2016). Due to these challenges, early crowding valuation studies relying on the RP data either deviated from DCMs (Kroes et al., 2014) or complemented the RP data with the SP data (Batarce et al., 2015). However, the emerging use of smart cards for fare collection provides an alternative way to collect the required RP data. Tirachini et al. (2016) first illustrated how smart card data can be used for the crowding valuation of the Mass Rapid Transit users in Singapore. Along the same lines, Hörcher et al. (2017) integrated the smart card data with the vehicle location data to estimate the crowding multiplier of Hong Kong Mass Transit Railway (MTR) users.

We identify three research gaps in the crowding valuation literature. First, whereas dynamic route preferences and learning behaviours are hard to capture in the SP experiments, previous RP studies also rely on static choice models. Second, a regular subway user might not actively make a compensatory route choice at every instance and can adhere to the same route until a bad experience occurs, but such non-compensatory behaviour has not been modelled by any of previous studies. Identification of such a non-compensatory choice process is crucial to avoid the overestimation of the crowding multiplier. Third, crowding on a subway route is not uniform, but it is assumed to be uniform in SP studies and aggregate measures are used in RP studies, without considering the spread and order of crowded journey lags. The sensitivity of travellers' route preferences (and thus, crowding valuation) relative to the crowding distribution over a route, i.e. the cost of crowding variability, remains an open question.

To address these gaps, we propose a dynamic latent class model (DLCM) which incorporates the learning behaviour of riders using the instance-based learning theory (IBLT) (Tang et al., 2017; Guevara et al., 2018), specifies compensatory and non-compensatory (i.e., inertia/habit) choice processes of subway users as latent classes, and allows users to dynamically transition between these classes based on the differences between the expected and the experienced level of services. Thus, the proposed DLCM provides a comprehensive and general framework to model dynamic choices. The resulting model turns out to be a new variant of the heterogeneous Hidden Markov Model (HMM) where a rider's choice at any instance not only depends on the rider's current state (i.e., latent class), but is also influenced by the rider's lagged choice. The proposed model also accounts for the unobserved heterogeneity in preferences of riders. We extend the expectation-maximization (EM) algorithm for HMMs to estimate the proposed DLCM and also adapt the Viterbi algorithm to predict the sequence of latent classes of a rider, conditional on her observed route choices (Arulampalam et al., 2002).

We apply the proposed DLCM to estimate the crowding cost of Hong Kong MTR riders using a four-month-long dynamic panel dataset on riders' revealed route preferences (Hörcher et al., 2017).

In doing so, we also explore whether crowding on the initial or the latter part of a trip is perceived more burdensome by subway passengers and whether the overall spread of crowding levels matters. This is the first such application in the crowding valuation literature.

We present experiment design in section 2, model formulation in section 3, estimation details in section 4, and a Monte Carlo study in section 5. Results of the empirical study will be provided in the full paper.

2 Implicit Experiment Design

We briefly discuss the RP experiment design (see Hörcher et al., 2017, for details). The selected network of Hong Kong MTR has 32 origin-destination (O-D) pairs with exactly two competitive paths between each O-D pair. These paths have enough variations in travel time and crowding, circumventing the concern of the dominant alternative. The routes chosen by passengers and attributes of routes are obtained by passing the day-to-day data on automated fare collection (AFC) and vehicle location (AVL) through our passenger-train assignment algorithm. The key attributes include travel time, the density of standing passengers, and the probability of standing. As fares are not differentiated based on the route chosen, we derive crowding cost valuations in terms of the equivalent travel time loss. Our dataset covers four months from two consecutive years, thus allowing for numerous repeated route choice observations from uniquely identified (but otherwise anonymised) smart card holders.

3 Model Formulation

The proposed DLCM has three components – initialisation model, transition model, and choice model. The long panel data allows us to utilize the first few observations of riders to identify their initial latent classes. We consider that a rider can choose to be in any of two latent classes (or hidden states) at a choice occasion: 1) compensatory, 2) non-compensatory, i.e. inclined to make choices due to habit or inertia. The rider's class transition probabilities depend on the difference between the expected and experienced level of service on the route chosen in the previous period. In the choice model, conditional on the rider's latent state and the lagged choice, the rider chooses a route from a set of two routes. In what follows, we provide a contextual description of the DLCM for two alternatives and two latent classes, but without loss of generality, it can be extended to any number of alternatives and latent classes. For simplicity, we first describe transition and initialisation models, followed by the choice model.

3.1 Transition Model

If rider i is in state s at time t, the utility M_{its} derived by her due to a mismatch between the expected and the experienced level of service at the chosen route j_t is:

$$M_{its} = m_{its} + \epsilon_{its} = \zeta_s^\top \left[\boldsymbol{X}_{itj_t} - \mathbb{E}(\boldsymbol{X}_{itj_t}) \right] + \epsilon_{its}, \tag{1}$$

where $\boldsymbol{X}_{itj_t}$ is a vector of attributes (i.e., travel time and crowding level) experienced by rider i on chosen route j_t at time t. We define expected values of these attributes $\mathbb{E}(\boldsymbol{X}_{itj_t})$ using IBLT (Tang et al., 2017):

$$\mathbb{E}(\boldsymbol{X}_{itj_t}) = \sum_{t'} W_{t't}\boldsymbol{X}_{it'j_t}, \quad t', \tau \in \{t-1, t-2, ...\},$$
$$\text{where} \quad W_{t't} = \frac{[t-t']^{-\mu}}{\sum_{\forall\tau}[t-\tau]^{-\mu}}, \tag{2}$$

where μ is a decay parameter that captures the rate of forgetting the past experiences. Assuming Gumbel distributed ϵ_{its}, transition probability expressions are:

$$P(s_{i(t+1)} = 1|s_{it} = s; \zeta_s, \mu) = \frac{\exp(m_{its})}{1+\exp(m_{its})},$$
$$P(s_{i(t+1)} = 2|s_{it} = s; \zeta_s, \mu) = 1 - P(s_{i(t+1)} = 1|s_{it} = s; \zeta_s, \mu), \tag{3}$$

If $\boldsymbol{X}$ includes the level-of-service attributes for which "less is better" (e.g., travel time, crowding), we expect $\boldsymbol{\zeta}_s$ to be positive. Intuitively, if the experienced level of service is poorer than that of the expected at time t, a rider is more likely to make a compensatory choice (state 1) at $t+1$.

3.2 Initialisation Model

Consider that we observe riders for $T_I + T$ periods. We use the first T_I choices of riders to infer their initial latent class probabilities. Similar to the transition model, we obtain the latent class probabilities of the rider after $t = T_I$ based on the differences between the experienced and expected level of service on the chosen route at $t = T_I$.

$$K_{iT_I} = k_{iT_I} + \epsilon_{iT_I} = \boldsymbol{\zeta}_0^\top \left[\boldsymbol{Z}_{iT_I j_{T_I}} - \mathbb{E}(\boldsymbol{Z}_{iT_I j_{T_I}})\right] + \epsilon_{iT_I},$$
$$P(s_{i(T_I+1)} = 1; \zeta_0, \mu) = \frac{\exp\left(k_{iT_I}\right)}{1+\exp\left(k_{iT_I}\right)}, \tag{4}$$
$$P(s_{i(T_I+1)} = 2; \zeta_0, \mu) = 1 - P(s_{i(T_I+1)} = 1; \zeta_0, \mu),$$

Note that the first $T_I - 1$ choices of the rider are mainly used to compute the expected level of service $\mathbb{E}(\boldsymbol{Z}_{iT_I j_{T_I}})$ for the route chosen by the rider at $t = T_I$. The expectation is computed using equation 2. If we shift the time clock by T_I periods, the latent class at $t = T_I + 1$, the choice at $t = T_I$, the choice at $t = T_I + 1$ correspond to the latent class at $t = 1$, the choice at $t = 0$, and the choice at $t = 1$, respectively. We do not include the first T_I choices of the rider in the choice model and thus, equation 4 provides the initial latent class probabilities $P(s_{i1}; \zeta_0, \mu)$.

3.3 Choice model

If rider i is in the compensatory state at time t (i.e., $s_{it} = 1$), her utility from choosing route j at time t is:

$$U_{itj} = V_{itj} + \nu_{itj} = \boldsymbol{\gamma}^\top \mathbb{E}(\boldsymbol{F}_{itj}) + \boldsymbol{\chi}_i^\top \mathbb{E}(\boldsymbol{G}_{itj}) + \nu_{itj}, \quad \text{where} \quad \boldsymbol{\chi}_i \sim \text{Normal}(\boldsymbol{\varrho}, \boldsymbol{\Psi}) \tag{5}$$

We consider that the marginal utility associated with attributes $\boldsymbol{F}_{itj}$ do not vary across riders, but preference heterogeneity is present for attributes $\boldsymbol{G}_{itj}$. The expected value of attributes is obtained using equation 2. If a rider is in the compensatory state at time t, she is less likely to choose the same route at time t as chosen at $t-1$. We can further account for this behaviour by modifying the indirect utility equation for the first route:

$$U_{it1} = V_{it1} + \lambda_1 \mathbb{1}[y_{i(t-1)} = 1] - \lambda_2 \mathbb{1}[y_{i(t-1)} = 2] + \nu_{it1} \tag{6}$$

where $\mathbb{1}[.]$ is an indicator function and the estimable parameters are $\boldsymbol{\Pi} = \{\boldsymbol{\gamma}, \boldsymbol{\varrho}, \boldsymbol{\Psi}, \mu, \lambda_1, \lambda_2\}$. Note that λ_1 and λ_2 are likely to be negative because they control the inherent aversion for choosing the same route. Considering Gumbel distribution on ν_{itj}, the route choice probabilities given that the passenger i is in the compensatory state at time t is:

$$\begin{aligned} P(y_{it} = 1 | s_{it} = 1, y_{i(t-1)}; \boldsymbol{\Pi}) &= \frac{\exp\left(V_{it1} + \lambda_1 \mathbb{1}[y_{i(t-1)} = 1] - \lambda_2 \mathbb{1}[y_{i(t-1)} = 2]\right)}{\exp\left(V_{it1} + \lambda_1 (\mathbb{1}[y_{i(t-1)} = 1] - \mathbb{1}[y_{i(t-1)} = 2])\right) + \exp\left(V_{it2}\right)} \\ P(y_{it} = 2 | s_{it} = 1, y_{i(t-1)}; \boldsymbol{\Pi}) &= 1 - P(y_{it} = 1 | s_{it} = 1, y_{i(t-1)}; \boldsymbol{\Pi}) \end{aligned} \tag{7}$$

where y_{it} is the route chosen by rider i at time t. Similarly, we now define the route choice probabilities if rider i is in non-compensatory state at t (i.e., $s_{it} = 2$):

$$\begin{aligned} P(y_{it} = 1 | s_{it} = 2, y_{i(t-1)}; \lambda_3, \lambda_4) &= \frac{\exp\left(\lambda_3 \mathbb{1}[y_{i(t-1)} = 1] - \lambda_4 \mathbb{1}[y_{i(t-1)} = 2]\right)}{\exp\left(\lambda_3 \mathbb{1}[y_{i(t-1)} = 1] - \lambda_4 \mathbb{1}[y_{i(t-1)} = 2]\right) + 1} \\ P(y_{it} = 2 | s_{it} = 2, y_{i(t-1)}; \lambda_3, \lambda_4) &= 1 - P(y_{it} = 1 | s_{it} = 2, y_{i(t-1)}; \lambda_3, \lambda_4) \end{aligned} \tag{8}$$

We would expect λ_3 and λ_4 to be highly positive because the passenger is likely to make the same choice in two consecutive scenarios due to inertia or habit.

4 Model Estimation

By combining all three components of the model and using the notations of HMMs, we write the conditional likelihood of the model:

$$\begin{aligned} P(y_{i1}, \ldots, y_{iT} | \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{F}, \boldsymbol{G}; \boldsymbol{\Theta}) = \sum_{s_1=1}^{2} \sum_{s_2=1}^{2} \cdots \sum_{s_T=1}^{2} \underbrace{\prod_{t=1}^{T} P\left(y_{it} | q_{its_t} = 1, y_{i(t-1)}\right)}_{\text{Choice Model}} \underbrace{P(q_{i1s_1} = 1 | \text{Inputs})}_{\text{Initialisation Model}} \ldots \\ \ldots \underbrace{\prod_{t=2}^{T} P(q_{its_t} = 1 | q_{i(t-1)s_{(t-1)}} = 1)}_{\text{Transition Model}} \end{aligned} \tag{9}$$

where q_{its} is 1 if the passenger i belongs to state s, else it is zero. The model parameters are $\boldsymbol{\Theta} = \left\{\mu, \zeta_0, \zeta_1, \zeta_2, \boldsymbol{\gamma}, \boldsymbol{\varrho}, \boldsymbol{\Psi}, \lambda_1, \lambda_2, \lambda_3, \lambda_4\right\}$. Figure 1 shows the schematic diagram of the proposed DLCM model. The proposed specification is a variant of the traditional heterogeneous hidden Markov models because conditional on the hidden state (i.e., latent class), choice probabilities also depend on the lagged choice.

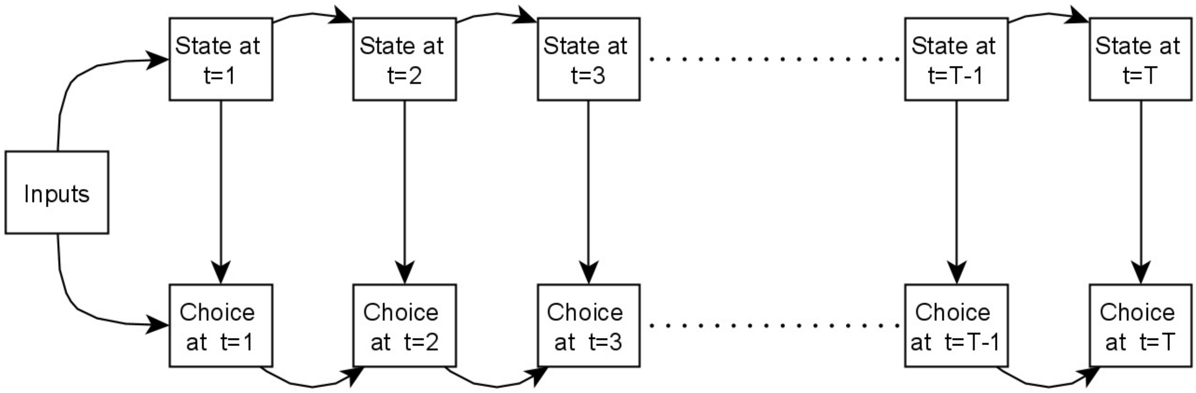

Figure 1: **The dynamic latent class model**

4.1 Expectation-Maximization (EM) Algorithm

The EM algorithm was originally developed to deal with the missing data problem. The DLCM likelihood maximisation problem also falls under the same category because latent classes can be treated as the missing data. The EM algorithm is a two-step iterative algorithm where the conditional expectation of the missing data is obtained in the E-step and then the complete loglikelihood is maximised in the M-step to update the model parameters. The convergence criterion is defined based on the difference in parameter estimates or loglikelihood values of two consecutive iterations.

Assuming latent classes as missing variables, we write the complete likelihood L_c and the complete loglikelihood $\mathscr{L}_c$ of the model:

$$\begin{aligned} L_c &= P(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T,\boldsymbol{s}_1,\ldots,\boldsymbol{s}_T;\boldsymbol{\Theta}) \\ &= \left[\prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{s=1}^{2}\left[P\left(y_{it}|q_{its}=1,y_{i(t-1)}\right)\right]^{q_{its}}\right]\left[\prod_{i=1}^{N}\prod_{s=1}^{2}[P(q_{i1s}=1|\text{Inputs})]^{q_{i1s}}\right]\ldots \\ &\ldots\left[\prod_{i=1}^{N}\prod_{t=2}^{T}\prod_{s=1}^{2}\prod_{r=1}^{2}[P(q_{its}=1|q_{i(t-1)r}=1)]^{q_{its}q_{i(t-1)r}}\right] \end{aligned} \tag{10}$$

$$\begin{aligned} \mathscr{L}_c = \log L_c &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{2}q_{its}\log\left[P\left(y_{it}|q_{its}=1,y_{i(t-1)}\right)\right]\right]\ldots \\ &\cdots+\left[\sum_{i=1}^{N}\sum_{s=1}^{2}q_{i1s}\log[P(q_{i1s}=1|\text{Inputs})]\right]\ldots \\ &\cdots+\left[\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{2}\sum_{r=1}^{2}[q_{its}q_{i(t-1)r}]\log\left[P(q_{its}=1|q_{i(t-1)r}=1)\right]\right] \end{aligned} \tag{11}$$

4.1.1 E-step

Based on the complete loglikelihood $\mathscr{L}_c$ expression, the E-step in $(k+1)^{th}$ iteration requires computing the following expectations:

$$\begin{aligned} \pi_{its}^{k+1} &= \mathbb{E}[q_{its}|\boldsymbol{y}_i;\boldsymbol{\Theta}^k] = P[q_{its}=1|\boldsymbol{y}_i;\boldsymbol{\Theta}^k] \\ \omega_{itrs}^{k+1} &= \mathbb{E}[q_{its}q_{i(t-1)r}|\boldsymbol{y}_i;\boldsymbol{\Theta}^k] = P[q_{its}q_{i(t-1)r}=1|\boldsymbol{y}_i;\boldsymbol{\Theta}^k] \end{aligned} \tag{12}$$

To compute expectations in the E-step efficiently, we define forward (α_{its}) and backward (β_{its}) variables:

$$\begin{aligned}\alpha_{its}(\mathbf{\Theta}) &= P(y_{i1}, \ldots, y_{it}, q_{its} = 1; \mathbf{\Theta}) \\ \beta_{its}(\mathbf{\Theta}) &= P(y_{i(t+1)}, \ldots, y_{iT} | y_{it}, q_{its} = 1; \mathbf{\Theta})\end{aligned} \tag{13}$$

We then compute the π_{its}^{k+1} and ω_{itrs}^{k+1} in terms of forward and backward variables using the Bayes theorem.

4.1.2 M-step

After computing π_{its}^{k} and ω_{itrs}^{k} in the E-step, the complete loglikelihood is maximised to obtain the parameters for $(k+1)^{\text{th}}$ iteration .

$$\begin{aligned}\mathbf{\Theta}^{k+1} = \arg\max_{\mathbf{\Theta}} \Bigg[& \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{2} \pi_{its}^{k} \log\left[P\left(y_{it} | q_{its} = 1, y_{i(t-1)}\right)\right] \\ & + \sum_{i=1}^{N}\sum_{s=1}^{2} \pi_{i1s}^{k} \log\left[P(q_{i1s} = 1 | \text{Inputs})\right] \\ & + \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s=1}^{2}\sum_{r=1}^{2} \omega_{itrs}^{k} \log\left[P(q_{its} = 1 | q_{i(t-1)r} = 1)\right] \Bigg]\end{aligned} \tag{14}$$

4.2 Sequence of Latent Classes

We use the Viterbi algorithm to estimate the most likely sequence of a rider's latent classes, conditional on the sequence of observed route choices. This algorithm uses forward-backward recursion (Arulampalam et al., 2002; Forney, 1973; He, 1988). Once we condition on the lagged choices in the recursion, the Viterbi algorithm for the heterogeneous HMMs can be used for the proposed DLCM.

5 Monte Carlo Study

To validate the model and properties of the EM estimator, we present an instance of a Monte Carlo study.[1]. We consider two data generating processes: i) with no preference heterogeneity; ii) with preference heterogeneity in the choice model. In both DGPs, we generate each component of explanatory variables $\{\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{F}, \boldsymbol{G}\}$ by taking draws from a normally-distributed random variable with mean 1.5 and standard deviation 0.3. We utilize the first ten choices (i.e., $T_I = 10$) of a rider to initialize the model and assume that the rider develops expectation for the level of service on a route based on her past three trips on that route. In both DGPs, we consider the memory parameter of the IBLT μ to be 1. We consider 3000 riders (i.e., $N = 3000$) and 120 observations per rider (i.e., $T + T_I = 120$) for the first DGP, but these values are 2000 and 30, respectively, for the second DGP. In the second DGP, we consider a diagonal variance-covariance matrix on random parameters. The algorithm terminates when the difference between the loglikelihood values of two consecutive iterations is below 10^{-6}.

[1]Since standard errors are calculated using the regular asymptotic theory, we don't repeat the Monte Carlo study on multiple resamples

Table 1: **Results of the Monte Carlo Study with Unobserved Heterogeneity**

	True value	Estimated value	Std. err.	t-stat	Gradient at convergence
Initialisation					
ζ_0^1	-1.0	-1.07	0.089	-12.0	2.77E-04
ζ_0^2	1.1	1.21	0.213	5.7	2.30E-05
ζ_0^3	0.9	0.84	0.204	4.1	2.70E-05
Transition Model (class 1)					
ζ_1^1	-1	-1.01	0.046	-21.7	6.40E-03
ζ_1^2	1.4	1.36	0.087	15.7	4.67E-04
ζ_1^3	1.5	1.33	0.085	15.6	4.44E-04
Transition Model (class 2)					
ζ_2^1	-1.5	-1.52	0.050	-30.5	1.71E-02
ζ_2^2	1.2	1.22	0.048	25.4	1.28E-03
ζ_2^3	1.1	1.13	0.045	25.3	1.02E-03
Choice Model (Class 1)					
λ_1	-1.0	-0.97	0.085	-11.5	4.35E-04
λ_2	-2.0	-2.02	0.173	-11.7	5.72E-04
γ^1	-1.0	-0.91	0.115	-7.9	4.53E-03
γ^2	-1.5	-1.59	0.152	-10.5	7.56E-03
Choice Model (Class 2)					
λ_3	1.0	0.99	0.023	43.2	9.05E-04
λ_4	2.0	1.98	0.055	35.8	3.02E-03

Table 2: **Results of the Monte Carlo Study with Unobserved Heterogeneity**

	True value	Estimated value	Std. err.	t-stat	Gradient at convergence
Initialisation					
ζ_0^1	-1.0	-1.08	0.14	-7.76	3.25E-03
ζ_0^2	1.1	1.34	0.27	4.90	1.30E-04
ζ_0^3	0.9	0.76	0.27	2.86	9.00E-05
Transition Model (class 1)					
ζ_1^1	-1.0	-1.05	0.14	-7.76	1.09E-02
ζ_1^2	1.4	1.68	0.34	4.95	8.27E-04
ζ_1^3	1.5	1.81	0.34	5.38	3.01E-04
Transition Model (class 2)					
ζ_2^1	-1.5	-1.63	0.13	-12.91	4.02E-02
ζ_2^2	1.2	1.22	0.13	9.35	1.48E-03
ζ_2^3	1.1	1.21	0.14	8.73	1.34E-03
Choice Model (Class 1)					
λ_1	-1.0	-0.84	0.24	-3.56	6.31E-03
λ_2	-2.0	-1.98	0.36	-5.52	7.48E-03
γ^1	-1.0	-1.05	0.32	-3.30	9.28E-04
γ^2	-1.5	-1.71	0.38	-4.55	8.66E-04
ϱ^1	1.5	1.46	0.27	5.36	-1.31E-03
ϱ^2	-1.5	-1.62	0.30	-5.35	1.66E-03
Ψ^{11}	1.0	0.89	0.77	1.15	1.42E-03
Ψ^{22}	1.0	1.25	0.70	1.78	1.91E-03
Choice Model (Class 2)					
λ_3	1.0	0.93	0.05	18.63	3.38E-03
λ_4	2.0	1.86	0.11	17.24	1.44E-02

Tables 1 and 2 present estimation results for both DGPs. In both tables, superscript on variable relates to the component number of the vector. For example, ζ_0^2 implies the second element of the vector ζ_0. A comparison of true and estimated values of parameters indicate that all model parameters are recovered well. Gradient values at convergence are also close to zero for all parameters in both DGPs, which ensure the convergence of the EM to a local optimal. Since true latent classes (or hidden states) of riders are known in the DGP, we could analyse the performance of the Viterbi algorithm in predicting hidden states. The results indicate that the Viterbi algorithm could predict latent states correctly at 84.2% and 84.8% accuracy in both DGPs, respectively.

References

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188.

Bansal, P., Hurtubia, R., Tirachini, A., and Daziano, R. A. (2019). Flexible estimates of heterogeneity in crowding valuation in the new york city subway. *Journal of Choice Modelling*, 31:124–140.

Batarce, M., Muñoz, J. C., de Dios Ortúzar, J., Raveau, S., Mojica, C., and Ríos, R. A. (2015). Use of mixed stated and revealed preference data for crowding valuation on public transport in santiago, chile. *Transportation Research Record*, 2535(1):73–78.

Forney, G. D. (1973). The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278.

Guevara, C. A., Tang, Y., and Gao, S. (2018). The initial condition problem with complete history dependency in learning models for travel choices. *Transportation Research Part B: Methodological*, 117:850–861.

He, Y. (1988). Extended viterbi algorithm for second order hidden markov process. In *[1988 Proceedings] 9th International Conference on Pattern Recognition*, pages 718–720. IEEE.

Hörcher, D., Graham, D. J., and Anderson, R. J. (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B: Methodological*, 95:105–125.

Kroes, E., Kouwenhoven, M., Debrincat, L., and Pauget, N. (2014). Value of crowding on public transport in île-de-france, france. *Transportation Research Record*, 2417(1):37–45.

Tang, Y., Gao, S., and Ben-Elia, E. (2017). An exploratory study of instance-based learning for route choice with random travel times. *Journal of choice modelling*, 24:22–35.

Tirachini, A., Sun, L., Erath, A., and Chakirov, A. (2016). Valuation of sitting and standing in metro trains using revealed preferences. *Transport Policy*, 47:94–104.

Wardman, M. and Whelan, G. (2011). Twenty years of rail crowding valuation studies: evidence and lessons from british experience. *Transport reviews*, 31(3):379–398.