Perturbed utility Markovian choice model: choice probability generation function and estimation

Rui Yao, Kenan Zhang[*]
School of Architecture, Civil and Environmental Engineering (ENAC),
École Polytechnique Fédérale de Lausanne (EFPL), Switzerland

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

Keywords: perturbed utility, Markov decision process, dynamic discrete choice, estimation

1 Introduction

Many choice problems in transportation can be modeled as a Markov decision process (MDP). One classic example is route choice constructed as a sequence of link choices. Specifically, at each node (*state*), the traveler chooses the next link (*action*) that maximizes a sum of the instantaneous random link utility (*reward*) and the expected maximum utility to the destination (*value function*). When the random fluctuation in link utility is additive and follows the generalized extreme value distribution (McFadden, 1981), the choice probability at each node has a closed-form expression, e.g., the recursive logit models (Fosgerau *et al.*, 2013, Mai, 2016, Oyama, 2023). Although the modeling framework is flexible, the existing estimation methods for these Markovian choice models all rely on a computationally demanding bi-level procedure (Rust, 1987): the upper level updates the parameter estimates, and the lower level solves the MDP problem using value iteration. Furthermore, when the test parameters are badly set, the lower level may fail to converge (Mai & Frejinger, 2022).

In this study, we propose a novel Markovian choice model based on the perturbed utility theory (Fosgerau & McFadden, 2012) and develop a highly efficient single-level estimation approach. At the core of the proposed model is a class of choice probability generation functions whose gradient directly maps from state-action value (Q-value) functions to a perturbed utility maximizing policy. Furthermore, the gradient mapping is invertible, a key property that helps reduce the complexity of model estimation. Remarkably, the estimation of any linear utility function requires only linear regression.

2 Perturbed utility Markovian choice model

2.1 Preliminaries

Perturbed utility discrete choice models (Fosgerau & McFadden, 2012) assume individuals decide on their choice probabilities to maximize a *perturbed utility* defined as the sum of the expected systematic utility and a convex perturbation function of the choice probabilities. Mathematically, the choice probabilities x are derived by solving

$$\max_{x \in B} \quad v^\top x - F(x), \tag{1}$$

[*] Corresponding author: kenan.zhang@epfl.ch

where v is the utility vector of alternatives, F denotes the convex perturbation function, and B is the feasible set of x.

The perturbed utility model (PUM) has been shown to generalize the additive random utility model (ARUM) (McFadden, 1981). For example, when the perturbation function is the Shannon entropy, the derived choice probabilities are equivalent to those in MNL. Despite its generality, determining the choice probabilities of PUM requires solving an optimization problem (1), which could be cumbersome when a large number of choices must be evaluated or the decision-making process has a recursive structure. Both of them, however, persist in the Markovian choice model. To tackle this challenge, we characterize a class of choice probability generation functions and establish conditions such that their gradient directly gives the optimal choice probabilities.

Let us first define the perturbed utility Markovian choice model. We consider a Markov decision process (MDP) with some termination state, thus the time horizon can be finite or infinite. The MDP is defined on a tuple $(\mathcal{S}, \mathcal{A}, P, u, \gamma)$, where $\mathcal{S}$ is the finite state space, $\mathcal{A}$ is the finite action space, $P : \mathcal{S} \times \mathcal{A} \to p(\mathcal{S})$ specifies state transition as the probability of transition between each pair of states under each action, $u \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ is the systematic utility, and $\gamma \in (0, 1]$ is the discount factor. For simplicity, we use $\mathcal{A}_s$ to denote the set of available actions at state $s \in \mathcal{S}$ and define $\Delta_s = \Delta(\mathcal{A}_s)$, the probability simplex of $\mathcal{A}_s$.

Following the common framework of MDP, we define value function $V : \mathcal{S} \to \mathbb{R}$ as the expected cumulative utility from a given state and define Q-value function as $Q(s, a) = u(s, a) + \gamma\mathbb{E}_{s'\sim P(\cdot|s,a)}[V(s')]$. We further define a state-dependent perturbation function F_s as a convex function of the conditional choice probability $\pi(\cdot|s) \in \Delta_s$, and assume $||\nabla F_s(\pi(\cdot|s))|| \to \infty$ as $\pi(\cdot|s)$ approaches the boundary of Δ_s. Such a property is also known as *essential smoothness* in the literature (Ch. 26 in Rockafellar, 1970) and it has been widely used in choice modeling (e.g., Fosgerau *et al.*, 2013). Accordingly, the conditional choice probability under PUMCM solves

$$\max_{\pi(\cdot|s)\in\operatorname{int}(\Delta_s)} \mathbb{E}_{a\sim\pi(\cdot|s)} [Q(s, a)] - F_s(\pi(\cdot|s)), \tag{2}$$

where $\operatorname{int}(\Delta_s)$ denotes the interior of Δ_s.

2.2 Choice probability generation function

In brief, a choice probability generation function H_s is a function of Q-values whose gradient gives the optimal conditional choice probabilities in PUMCM. Therefore, we can bypass solving (2) and directly obtain the choice probabilities when Q-values are known. The general conditions for a choice probability generation function are formally stated in the following proposition:

Proposition 1 *Suppose a function $H_s : \mathbb{R}^{|\mathcal{A}_s|} \to \mathbb{R}$ defined on a state $s \in \mathcal{S}$ satisfies: i) twice continuously differentiable, ii) gradient falls in the interior of simplex Δ_s, i.e., $\nabla H_s(Q(s, \cdot)) \in \operatorname{int}(\Delta_s), \forall Q(s, \cdot) \in \mathbb{R}^{|\mathcal{A}_s|}$, and iii) Hessian matrix $\nabla^2 H_s(Q(s, \cdot))$ is positive definite on T_s for all $Q(s, \cdot) \in \mathbb{R}^{|\mathcal{A}_s|}$, where $T_s := \left\{z \in \mathbb{R}^{|\mathcal{A}_s|} | \sum_j z_j = 0\right\}$ denotes the tangent space of Δ_s. Then, there exists a convex perturbation function $H^*_s : \operatorname{int}(\Delta_s) \to \mathbb{R}$, such that the gradient of H_s at $Q(s, \cdot)$ solves the perturbed utility maximization problem:*

$$\nabla H_s(Q(s, \cdot)) = \arg \max_{\pi(\cdot|s)\in\operatorname{int}(\Delta_s)} \mathbb{E}_{a\sim\pi(\cdot|s)} [Q(s, a)] - H^*_s(\pi(\cdot|s)). \tag{3}$$

In other words, $\nabla H_s(Q(s, \cdot))$ gives the choice probabilities in PUMCM. Moreover, ∇H_s is invertible on T_s and $(\nabla H_s)^{-1} \equiv \nabla H^*_s : \operatorname{int}(\Delta_s) \to T_s$.

The following corollary describes a particular property that enables efficient estimation:

Corollary 1 *Suppose H_s satisfies the conditions listed in Proposition 1. Then, for any $Q(s, \cdot) \in \mathbb{R}^{|\mathcal{A}_s|}$, there exists a constant $K_s \in \mathbb{R}$ and $Q_0(s, \cdot) \in T_s$ such that*

$$Q(s, \cdot) = Q_0(s, \cdot) + K_s\mathbf{1}, \tag{4}$$

$$H_s(Q(s, \cdot)) = H_s(Q_0(s, \cdot)) + K_s. \tag{5}$$

2.3 Model estimation

We now proceed to discuss the estimation of a parametric utility function in PUMCM. Suppose the observed choices follow the optimal conditional choice probabilities π^*. To begin with, we rewrite the optimal Q-value in the matrix form:

$$Q^* = u + \gamma P V^*, \tag{6}$$

where $u = (u(s,a))^\top_{s\in\mathcal{S},a\in\mathcal{A}_s} \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}_s|}$ is the utility vector, $P = (P(\cdot|s,a))^\top_{s\in\mathcal{S},a\in\mathcal{A}_s} \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}_s|\times|\mathcal{S}|}$ is the transition matrix, and $V^* = (V^*(s))^\top_{s\in\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$ is the vector of optimal values.

Then for each $s \in \mathcal{S}$, we have

$$Q^*(s,\cdot) = Q^*_0(s,\cdot) + K_s\mathbf{1} = (\nabla H_s)^{-1}(\pi^*(\cdot|s)) + K_s\mathbf{1} = \nabla H^*_s(\pi^*(\cdot|s)) + K_s\mathbf{1}. \tag{7}$$

The first equality directly applies the result in Corollary 1, the second evokes the invertibility of ∇H_s on the tangent space T_s derived in Proposition 1, and the third replaces $(\nabla H_s)^{-1}$ with its corresponding perturbation function, another result of Proposition 1. In other words, for any observed π^*, the optimal Q-values are known up to a constant K_s. The following proposition further demonstrates the optimal values can be revealed from π^* under mild assumptions.

Proposition 2 *Suppose the feasible set of values, $\mathcal{M}$, is compact. Then, for each $s \in \mathcal{S}$, there exists unique $V^*(s)$ such that $V^*(s) = H_s(Q^*(s,\cdot))$.*

We note that the compactness of $\mathcal{M}$ implies V^* is bounded, which naturally holds in MDP with a finite horizon or with a discounted infinite horizon ($\gamma < 1$). It is also a reasonable assumption for undiscounted infinite horizon problems ($\gamma = 1$) with termination states (e.g., the destination in route choice).

Combining all the above analytical results, we have for each $s \in \mathcal{S}$,

$$V^*(s) = H_s(Q^*(s,\cdot)) = H_s(Q^*_0(s,\cdot)) + K_s = H_s(\nabla H^*_s(\pi^*(\cdot|s))) + K_s. \tag{8}$$

Let $\mathcal{Q} = (\nabla H^*_s(\pi^*(a|s)))^\top_{s\in\mathcal{S},a\in\mathcal{A}_s} \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}_s|}$ and $\mathcal{V} = (H_s(\nabla H^*_s(\pi^*(\cdot|s))))^\top_{s\in\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$, and $\Lambda \in \{0,1\}^{|\mathcal{S}||\mathcal{A}_s|\times|\mathcal{S}|}$, where $\Lambda_{(s,a),s} = 1, \forall s \in \mathcal{S}, a \in \mathcal{A}_s$, and zero, otherwise. Plugging Eqs. (8) and (7) with their matrix forms into Eq. (6) yields

$$Q^* = \mathcal{Q} + \Lambda K = u + \gamma P(\mathcal{V} + K) \;\Rightarrow\; \mathcal{Q} - \gamma P\mathcal{V} = u + (\gamma P - \Lambda)K = u + \mathcal{P}K, \tag{9}$$

where $\mathcal{P} = \gamma P - \Lambda$.

We are now ready to formulate the model estimation problem. With the observed policy π^*, the presumed generation function H_s and its corresponding perturbation function ∇H^*_s, we first derived $\mathcal{Q}$ and $\mathcal{V}$, then compute $\mathcal{Y} = \mathcal{Q} - \gamma P\mathcal{V}$. Let the utility function $u(Z,\beta)$ defined on attributes Z and parameter β, then Eq. (9) is rewritten as

$$\mathcal{Y} = u(Z,\beta) + \mathcal{P}K. \tag{10}$$

Eq. (10) can be simplified as $J\mathcal{Y} = Ju(Z,\beta)$ by introducing a projection matrix $J = B - (\mathcal{P}^\top B)^+\mathcal{P}^\top B$, where $B = \operatorname{diag}(1_{\pi>0})$ and $(\cdot)^+$ denotes the Moore-Penrose inverse (Fosgerau *et al.*, 2022). In this way, the constant K is eliminated and the problem further reduces to a linear regression when the utility function is linear, i.e., $u(Z,\beta) = Z\beta$.

3 Simulation experiment

We demonstrate the proposed PUMCM and its estimation using a simple route choice problem on a 13×13 bidirectional grid network. The state and action spaces correspond to the node and link sets, respectively, and the state transition is accordingly the link-node incident matrix. We consider a linear link utility function $u(Z, \beta) = Z\beta$, where the true values of β are reported in Table 1 and attributes Z are uniformly sampled between $[15, 45]$. Finally, the discount factor is set to $\gamma = 1$ following the literature on route choice.

A synthetic dataset of route choices is generated by performing random walks from 1000 randomly selected origins to a single destination. We consider the choice probability generation function $H_s(Q(s, \cdot)) = \ln(\sum \exp(Q(s, \cdot)))$, which leads to the recursive logit model (Fosgerau *et al.*, 2013, Mai, 2016, Oyama, 2023), and solve the optimal routing policies via value iterations.

Table 1 – *Mean and stdev. (in brackets) of parameter estimates over 10 replications.*

	β_1	β_2	β_3
True β	-0.0500	-0.1000	0.0500
$\hat{\beta}, \forall\, u^* \leq 0$	-0.0496 (0.0040)	-0.0990 (0.0031)	0.0480 (0.0020)
$\hat{\beta}, \exists\, u^* > 0$	-0.0482 (0.0011)	-0.0937 (0.0034)	0.0439 (0.0026)

We consider two scenarios: a default scenario where all link utilities are non-positive (non-negative travel costs), and a less common scenario where some links have positive utilities but the optimal values are bounded (no infinite loop). As shown in Table 1, the parameter estimates $\hat{\beta}$ are close to the true values in both scenarios.

4 Conclusion

This paper proposes the perturbed utility Markovian choice model (PUMCM), where sequential decisions are modeled as a Markov decision process and maximize a perturbed utility at each state. A class of choice probability generation functions is characterized, whose gradient is the optimal policy. An efficient estimation approach is then developed and demonstrated via numerical experiments. To the best of our knowledge, both PUMCM and its estimation are novel and complement their static counterpart (Fosgerau *et al.*, 2022, Yao *et al.*, 2024).

References

Fosgerau, Mogens, & McFadden, DL. 2012. A theory of the perturbed consumer with general budgets. *NBER Working Paper*, **17953**.

Fosgerau, Mogens, Frejinger, Emma, & Karlstrom, Anders. 2013. A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, **56**, 70–80.

Fosgerau, Mogens, Paulsen, Mads, & Rasmussen, Thomas Kjær. 2022. A perturbed utility route choice model. *Transportation Research Part C: Emerging Technologies*, **136**, 103514.

Mai, Tien. 2016. A method of integrating correlation structures for a generalized recursive route choice model. *Transportation Research Part B: Methodological*, **93**, 146–161.

Mai, Tien, & Frejinger, Emma. 2022. Undiscounted recursive path choice models: Convergence properties and algorithms. *Transportation Science*, **56**(6), 1469–1482.

McFadden, D. 1981. *Econometric Models of Probabilistic Choice".*

Oyama, Yuki. 2023. Capturing positive network attributes during the estimation of recursive logit models: A prism-based approach. *Transportation Research Part C: Emerging Technologies*, **147**, 104014.

Rockafellar, Ralph Tyrell. 1970. *Convex Analysis*. Princeton: Princeton University Press.

Rust, John. 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

Yao, Rui, Fosgerau, Mogens, Paulsen, Mads, & Rasmussen, Thomas Kjær. 2024. Perturbed utility stochastic traffic assignment. *Transportation Science*.