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

Rui Yao[1] and Kenan Zhang*[1]

[1]School of Architecture, Civil and Environmental Engineering (ENAC),École Polytechnique Fédérale de Lausanne (EFPL), Switzerland

Short summary

This paper proposes the perturbed utility Markovian choice model (PUMCM), where sequential decisions of individuals are modeled as a Markov decision process that maximizes a perturbed utility at each state. A class of choice probability generation functions is characterized, whose gradient directly yields the optimal policy. An efficient single-level estimation approach is then developed by leveraging the invertibility of the gradient mapping of the choice probability generation function. Notably, the proposed estimation method eliminates the need for the computationally intensive bi-level estimation that is commonly used in existing Markovian choice models. Further, our approach is robust in the sense that it allows both positive and negative parameters, which is demonstrated through numerical experiments. To the best of our knowledge, both PUMCM and its estimation are novel and complement to their static counterpart of perturbed utility-based choice models.
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 (GEV) distribution (McFadden, 1981), the choice probability at each node has a closed-form expression, e.g., the recursive logit, recursive nested logit, recursive network GEV 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, especially for cases beyond the recursive logit model (Mai & Frejinger, 2022). To mitigate the convergence issue, the undiscounted and infinite-horizon MDP can be restricted to a discounted one (Oyama & Hato, 2017), or a finite-horizon one (Oyama, 2023), but both at the cost of generality.

In this study, we propose a novel Markovian choice model based on the perturbed utility theory (Fosgerau & McFadden, 2012; Hofbauer & Sandholm, 2002) 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, avoiding the need to explicitly solve for the optimal policy. Furthermore, we established that such gradient mapping is invertible on a subspace, a key property that great reduces the complexity of model estimation. Remarkably, the estimation of any linear utility function requires only linear regression.

In what follows, we will first briefly review the perturbed utility theory, then continue with the formal definitions of perturbed utility Markovian choice model (PUMCM) and choice probability generation functions, discuss model estimation, and end with a discussion on numerical results.

*Corresponding author: kenan.zhang@epfl.ch

2 Perturbed utility Markovian choice model

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

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

where v is the utility vector of alternatives, F denotes the essentially 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 multinomial lofit model (MNL). Despite its generality, determining the choice probabilities of PUM often 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

$$Q(s, a) = u(s, a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[V(s')] \tag{2}$$

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., Hofbauer & Sandholm, 2002). For instance, Shannon entropy is essentially smooth, and the resulting choice probabilities are at the interior of the probability simplex (Fosgerau et al., 2013). Accordingly, the conditional choice probability under PUMCM solves

$$\max_{\pi(\cdot|s) \in \text{int}(\Delta_s)} \mathbb{E}_{a \sim \pi(\cdot|s)} \left[Q(s, a) \right] - F_s(\pi(\cdot|s)), \tag{3}$$

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

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 (3) 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:*

1. *twice continuously differentiable,*
2. *gradient falls in the interior of simplex Δ_s, i.e., $\nabla H_s(Q(s, \cdot)) \in \text{int}(\Delta_s), \forall Q(s, \cdot) \in \mathbb{R}^{|\mathcal{A}_s|}$,*
3. *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 $H_s^* : \text{int}(\Delta_s) \to \mathbb{R}$, such that the gradient of H_s at $Q(s,\cdot)$ solves the perturbed utility maximization problem:

$$\nabla H_s(Q(s,\cdot)) = \arg \max_{\pi(\cdot|s)\in\text{int}(\Delta_s)} \mathbb{E}_{a\sim\pi(\cdot|s)}\left[Q(s,a)\right] - H_s^*(\pi(\cdot|s)). \tag{4}$$

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

The proof is based on Legendre transformation (Rockafellar, 1970; Boyd & Vandenberghe, 2004) and will be provided in the full paper. We note that several well-known functions satisfy these conditions. For instance, the surplus functions of ARUM, and specifically, the log-sum-exp function of the MNL model. We will provide choice probability generation functions H_s and their corresponding perturbation functions H_s^* for the recursive nested logit and the more general network GEV model in the full paper.
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*

$$Q(s,\cdot) = Q_0(s,\cdot) + K_s\mathbf{1}, \tag{5}$$
$$H_s(Q(s,\cdot)) = H_s(Q_0(s,\cdot)) + K_s. \tag{6}$$

Corollary 1 essentially states that, for the class of choice probability generation function according to Proposition 1, there is exactly one degree of freedom as captured by the constant K_s. Building upon this observation, we develop the single-level estimation method.

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:

$$Q^* = u + \gamma P V^*, \tag{7}$$

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

$$Q^*(s,\cdot) = Q_0^*(s,\cdot) + K_s\mathbf{1} = (\nabla H_s)^{-1}(\pi^*(\cdot|s)) + K_s\mathbf{1} = \nabla H_s^*(\pi^*(\cdot|s)) + K_s\mathbf{1}. \tag{8}$$

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. As a result, we connect Q_0^* with observations π^*. Then, the third equality replaces $(\nabla H_s)^{-1}$ with its corresponding perturbation function, another result of Proposition 1. Hence, 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, there exists unique $V^* \in \mathcal{M}$ such that $V^*(s) = H_s(Q^*(s,\cdot))$ for each $s \in \mathcal{S}$.*

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 $s \in \mathcal{S}$,

$$V^*(s) = H_s(Q^*(s,\cdot)) = H_s(Q_0^*(s,\cdot)) + K_s = H_s(\nabla H_s^*(\pi^*(\cdot|s))) + K_s. \tag{9}$$

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. (9) and (8) with their matrix forms into Eq. (7) yields

$$Q^* = \mathcal{Q} + \Lambda K = u + \gamma P(\mathcal{V} + K) \;\Rightarrow\; \mathcal{Q} - \gamma P\mathcal{V} = u + (\gamma P - \Lambda)K = u + \mathcal{P}K, \tag{10}$$

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 (Jacobian of) 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. (10) is rewritten as

$$\mathcal{Y} = u(Z, \beta) + \mathcal{P}K. \tag{11}$$

Eq. (11) 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., $u(Z, \beta) = Z\beta$, such that the parameters β have closed-form:

$$\beta = [(JZ)^\top (JZ)]^{-1} (JZ)^\top J\mathcal{Y}, \tag{12}$$

where invertibility follows the typical full rank condition on data matrix JZ for linear regression.

3 Simulation experiment and discussion

We demonstrate the proposed PUMCM and its estimation using a simple route choice problem on a 13×13 bidirectional grid network consisting of 169 nodes and 624 links. The state and action spaces correspond to the node set $\mathcal{N}$ and link set $\mathcal{E}$, respectively, and the state transition is accordingly the node-link incident matrix $\mathcal{A} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{E}|}$ has entries

$$a_{v,ij} = \begin{cases} -1, & v = i, \\ 1, & v = j, \\ 0, & \text{otherwise.} \end{cases}$$

We consider a linear link utility function

$$u(Z, \beta) = Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3$$

where the true values of $\beta = (\beta_1, \beta_2, \beta_3)$ are reported in Table 1 and attributes $Z = (Z_1, Z_2, Z_3)$ are independently and uniformly sampled between an arbitrarily chosen range $[15, 45]$. 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 (with replacement) to a single destination that is also randomly selected. We consider an entropy-based choice probability generation function and its corresponding perturbation function as

$$H_s(Q(s, \cdot)) = \ln(\sum \exp(Q(s, \cdot))),$$
$$H_s^*(\pi(\cdot|s)) = \pi(\cdot|s)[\ln(\pi(\cdot|s)) - 1],$$

which correspond to the recursive logit model (Fosgerau et al., 2013; Mai, 2016; Oyama, 2023). To simulate route choices, we need first to compute the *true* optimal (routing) policy π^* under the preset β. As shown in Proposition 1, the optimal policy can be computed using $\pi^*(\cdot|s) = \nabla H_s(Q^*(s, \cdot))$ with $Q^*(s, \cdot)$ solved via value iterations. With the simulated route choices as random walks under policy π^*, we then compute the observed link choice frequencies at each node. This procedure mimics the data collection and processing in real applications: route choice observations are aggregated into link choice frequencies as the inputs for parameter estimation. As the number of observations increases, the link choice frequencies shall approach the true policy π^*.

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	-0.0990	0.0480
	(0.0040)	(0.0031)	(0.0020)
$\hat{\beta}, \exists\, u > 0$	-0.0482	-0.0937	0.0439
	(0.0011)	(0.0034)	(0.0026)

In most previous studies on recursive logit model (e.g., Fosgerau et al., 2013), parameters are restricted to be negative to ensure convergence of the existing bi-level estimation procedure. Such a constraint ensures bounded optimal values so that the value iteration to solve the lower-level problem (i.e., optimal route choice given a set of parameters) can always converge. It clearly leads to bias when some parameters are essentially non-negative. Alternatively, Mai & Frejinger (2022) suggest using second-order algorithms (e.g., Newton-based methods) to solve the lower-level problem, which does no longer require parameters to be negative. Although providing better empirical convergence performance, this approach still lacks a theoretical convergence guarantee when the candidate parameters are badly selected. Another way to ensure convergence is to turn the undiscounted infinite-horizon MDP into either discounted Oyama & Hato (2017) or finite-horizon (Oyama, 2023), which, however, loses the behavioral interpretation.

In contrast, our proposed single-level estimation method avoids solving the optimal values but directly obtains parameter estimates via regression. Therefore, either positive or negative parameters can be recovered, so long as the *true* optimal values are bounded as per Proposition 2. Note that this condition is far more relaxed than aforementioned studies, which requires the optimal values under *all candidate parameters* are bounded. Besides, bounded optimal values naturally hold in reality because they imply that there is no infinite loop. In other words, travelers would keep routing in the network without reaching their destinations.

To demonstrate this particular advantage of our proposed estimation method, we construct 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. In the latter, we ensure that no circle has a positive utility, which in turn, guarantees all optimal values are bounded. As shown in Table 1, the parameter estimates $\hat{\beta}$ are close to the true values in both scenarios, though the estimates in the first scenario are closer to the true values. A closer look into the route choice observations reveals that more links are visited in the first scenario (592/624) compared to the second (549/624). If we count the median of observations on each link, the first scenario (11 per link) is also higher than the second (7 per link). This difference is due to the existence of positive link utilities in the second scenario, which make some links very "attractive" and thus concentrate more choices. Consequently, data collected in the first scenario shows a greater variation than the second, which possibly results in its higher accuracy of parameter estimates.

4 Conclusion

In this study, we introduced the Perturbed Utility Markovian Choice Model (PUMCM), a novel framework for modeling sequential decision-making processes as Markov decision processes (MDPs) that maximize perturbed utility. One key innovation of this model lies in the characterization of choice probability generation functions, whose gradients directly yield the optimal policy.

We developed an efficient single-level estimation method that leverages the invertibility of the gradient mapping of these generation functions, significantly reducing the complexity of parameter estimation. Notably, the proposed estimation method eliminates the need for computationally intensive bi-level estimation procedures commonly used in existing Markovian choice models. Further, our approach allows for the recovery of both positive and negative parameters, overcoming the limitation of previous methods relying on restrictive assumptions to ensure convergence.

Through numerical experiments on a hypothetical route choice problem, we demonstrated the robustness and flexibility of the PUMCM. Particularly, we show that the proposed method can accurately estimate parameters even in scenarios where some link utilities are positive, provided that the optimal values remain bounded. To the best of our knowledge, both PUMCM and its estimation are novel and complement to their static counterpart of perturbed utility-based choice models (Fosgerau et al., 2022; Yao et al., 2024).

Future research could explore the application of PUMCM to more complex decision-making scenarios, such as activity-based models. Additionally, the integration of real-world data could further validate the model's practical utility and enhance its applicability in fields such as transportation planning, logistics, and beyond.

References

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

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

Fosgerau, M., & McFadden, D. (2012). A theory of the perturbed consumer with general budgets. *NBER Working Paper*, *17953*.

Fosgerau, M., Paulsen, M., & Rasmussen, T. K. (2022). A perturbed utility route choice model. *Transportation Research Part C: Emerging Technologies*, *136*, 103514.

Hofbauer, J., & Sandholm, W. H. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, *70*(6), 2265–2294.

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

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

McFadden, D. (1981). *Econometric models of probabilistic choice".* Structural Analysis of Discrete Data with Econometric Applications/The MIT Press.

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

Oyama, Y., & Hato, E. (2017). A discounted recursive logit model for dynamic gridlock network analysis. *Transportation Research Part C: Emerging Technologies*, *85*, 509–527.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton: Princeton University Press.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

Yao, R., Fosgerau, M., Paulsen, M., & Rasmussen, T. K. (2024). Perturbed utility stochastic traffic assignment. *Transportation Science*.