A perturbed utility model with semi-nonparametric perturbation function

Rui Yao[a], Xuhang Liu[a], Mogens Fosgerau[b], and Kenan Zhang[a,*]

[a] EPFL, Lausanne, Switzerland

[b] University of Copenhagen, and Technical University of Denmark, Copenhagen, Denmark

rui.yao@epfl.ch, xuhang.liu@epfl.ch, mogens.fosgerau@econ.ku.dk, kenan.zhang@epfl.ch

[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)

September 02-04, 2024, Crete, Greece

Keywords: perturbed utility model, discrete choice, semi-nonparametric, model identification

1 Introduction

In general perturbed utility discrete choice models (Fosgerau & McFadden, 2012, Allen & Rehbeck, 2019), each individual's decision is modeled as probabilities of choosing discrete alternatives and the individual is assumed to optimize choice probability to maximize a so-called *perturbed utility*, which is defined as the sum of the expected systematic utility and a convex perturbation function of the choice probabilities. Formally, it assumes each individual n solves the following maximization problem:

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

where x_n denote the choice probabilities of n across alternatives, v is the vector of systematic utilities, F denote the convex perturbation function, and B_n is the feasible set of x_n. For the notation simplicity, we drop the index of individual n hereafter.

The perturbed utility model (PUM) has been shown to generalize the additive random utility model discrete choice model (McFadden, 1981) as well as a range of other models. For example, when the perturbation function is the Shannon entropy, the derived choice probabilities have the form of the multinomial logit model (MNL). In spite of its generality, the general PUM allows identification of the systematic utility components. In particular, parameters in the systematic utility can be uniquely estimated up to normalization, provided the perturbation function is convex (Allen & Rehbeck, 2019).

However, to predict the choice probability of individuals, we must also specify a proper perturbation function, which is extremely challenging because there is no prior knowledge of its functional form. Existing studies often resort to predefined simple forms (Fosgerau *et al.*, 2024), but the resulting misspecification of the perturbation function is prone to sacrifice model fit. On the other hand, fully nonparametric forms maintain the flexibility required by the perturbation function but are typically hard to estimate (Chen, 2007). In this paper, we propose a *semi-nonparametric* form for the perturbation function that retains its generality meanwhile guaranteeing estimation performance with finite samples. To the best of our knowledge, the semi-nonparametric approach to the modeling and estimation of perturbation function in PUM has not been studied in the literature. We establish the identifiability of the resulting PUM, develop an efficient estimation approach, and demonstrate its performance through Monte Carlo simulations.

2 A perturbed utility model with semi-nonparametric perturbation function

We consider a perturbed utility model (PUM) with linear choice probability constraints as follows:

$$\max_{x \in \mathbb{R}_+^E} \quad v^\top x - w^\top F(x), \tag{2a}$$

$$\text{s.t.} \quad Ax = b, \tag{2b}$$

where $x = (x_1, ..., x_e, ..., x_E)$ is the choice probability vector of dimension E, and the total perturbation is defined as the sum over component-wise perturbations $F(x) = (F(x_1), ..., F(x_e), ..., F(x_E))$ with exogenous weights $w = (w_1, ..., w_e, ..., w_E)$.

Our proposed perturbation function $F(\cdot)$ for each component is represented by the sum of R integrals as follows:

$$F(x_e) = \sum_{r=1}^{R} \mu_r \int_0^{x_e} [\sigma(\alpha_r u + \gamma_r) - \sigma(\gamma_r)] \mathrm{d}u, \tag{3}$$

where $\sigma(\cdot)$ is a sigmoidal function, and $\mu_r, \alpha_r, \gamma_r$ are parameters that satisfy $\mu_r > 0, \alpha_r > 0, \forall r$.

Perturbation (3) is considered semi-nonparametric when R increases with the sample size, and can approximate the unknown function arbitrarily well (Chen, 2007). Besides, one can easily verify that the perturbation function (3) is strictly convex as its second derivative

$$\nabla^2 F(x_e) = \sum_{r=1}^{R} \mu_r \alpha_r \sigma'(\alpha_r x_e + \gamma_r) > 0, \tag{4}$$

where $\sigma'(\cdot)$ is the derivative of sigmoidal function and is strictly positive.

Further, Model (2) generalizes a large variety of choice scenarios in transportation systems. For instance, when A is the node-link incident matrix and b is the unit demand vector, the resulting model represents the network route choice problem (Fosgerau *et al.*, 2022).

2.1 Identification

Let $v_e : \mathbb{R}^D \to \mathbb{R}$ be the systematic utility function of alternative e of attribute vector $z_e \in \mathbb{R}^D$. Then, we define $v(z) = (v_1(z_1), ..., v_e(z_e), ..., v_E(z_E))$, the vector of systematic utilities, as a function of alternative attribute matrix $z = (z_1, ..., z_e, ..., z_E)^\top \in \mathbb{R}^{E \times D}$. Accordingly, the identification problem of Model (2) regards estimating the parameters in both $v(\cdot)$ and $F(\cdot)$. In what follows, we present the key results upon which we establish the identifiability of Model (2) with the semi-nonparametric perturbation function (3).

Our first result regards the identification of the perturbation value. Let $x_e^*(v(z))$ denote the optimal solution of x_e for a given z. The following lemma, adopted from Corollary 1 in Allen & Rehbeck (2019), gives the conditions for $F(x_e^*(v(z))), \forall e$ to be uniquely determined.

Lemma 1 *Suppose i) $v(\cdot)$ is known, convex and everywhere finite, and ii) $F(\cdot)$ is defined as per Eq.* (3) *and everywhere finite. Then, for every $z^1 \in supp(z)$, $F(x_e^*(v(z^1)))$ is uniquely determined if there exists $z^0 \in supp(z)$ such that $x_e^*(v(z^0)) = 0$.*

Secondly, we prove the parameters in $F(\cdot)$ are identifiable. To this end, we need to introduce the notion of functional equivalence (Albertini *et al.*, 1993) as follows:

Definition 1 (Equivalent perturbation functional) *Let $F, \tilde{F}$ be two perturbation functions defined by Eq.(3), we say F is functional equivalent to $\tilde{F}$, if 1) $R = \tilde{R}$, equal number of sigmoidal components; 2) (μ, α, γ) can be transformed into $(\tilde{\mu}, \tilde{\alpha}, \tilde{\gamma})$ in finite number of interchanging $(\mu_r, \alpha_r, \gamma_r)$ with $(\mu_{r'}, \alpha_{r'}, \gamma_{r'})$ and/or simultaneous sign-flipping of (μ_r, α_r).*

We then derive the identification conditions for parameters in $F(\cdot)$ by evoking Lemma 2.2 in Albertini *et al.* (1993).

Lemma 2 *Suppose i) $F(x)$ is uniquely determined $\forall x \in [0, 1]$, and ii) $F(\cdot)$ satisfies the no-clone condition, i.e., $(\alpha_r, \gamma_r) \neq \pm(\alpha_{r'}, \gamma_{r'}), \forall r \neq r'$. Then, parameters (R, μ, α, γ) are identifiable as per functional equivalence.*

With Lemmas 1 and 2, we finally establish the identifiability of our proposed PUM.

Theorem 1 (Model identification) *Under the assumptions of Lemmas 1 and 2, Model* (2) *with perturbation function* (3) *is identifiable in the sense that:*

- *The systematic utility $v(\cdot)$ is identifiable up to normalization, i.e., i) $\partial v_e(z_e)/\partial z_{e,d} \in \{1, -1\}$ for some attribute d in alternative e; and ii) $v_e(\mathbf{0}) = 0, \forall e$.*
- *The perturbation $F(\cdot)$ is identifiable as per functional equivalence.*

All the proofs will be included in the full paper.

2.2 Estimation

We develop a least-square estimator for the proposed PUM that is inspired by the *method of sieve* proposed by Grenander (1981) and the projection matrix introduced in Fosgerau *et al.* (2022). To start with, we impose a linear structure of the systematic utility as $v(z) = z\beta$ with parameter $\beta \in \mathbb{R}^D$ (i.e., $v_e(z_e) = z_e^\top \beta$). Accordingly, the Lagrangian of Model (2) is given by

$$L(x, \eta) = (z\beta)^\top x - w^\top F(x) + \eta^\top (Ax - b), \ x \in \mathbb{R}_+^E, \tag{5}$$

where η are the dual variables for constraints (2b). Let $B = \text{diag}(1_{x>0})$ be the matrix with ones on the diagonal for positive x_e, then the first-order condition can be derived as

$$B\left(z\beta - w \circ \nabla F(x) + A^\top \eta\right) = 0, \tag{6}$$

where $\circ$ denotes element-wise multiplication and $\nabla F(x) = (\nabla F(x_1), ..., \nabla F(x_e), ..., \nabla F(x_E))$ is the gradient vector of $F(x)$. To eliminate the dual variables, we introduce the projection matrix

$$P = B - (AB)^+ AB, \tag{7}$$

where $(AB)^+$ denotes the Moore-Penrose inverse of AB. Pre-multiplying Eq. (6) by P then yields the projected first-order condition

$$P(z\beta - w \circ \nabla F(x)) = 0. \tag{8}$$

Denote $\theta = (\beta, \mu, \alpha, \gamma)$. A fixed point is thus constructed using Eq. (8) as

$$x = x - P(z\beta - w \circ \nabla F(x)) = \Psi(x, \theta) \tag{9}$$

Accordingly, a least-square estimation problem with N samples is defined as

$$\min_\theta \frac{1}{N} \sum_{n=1}^{N} ||x_n - \Psi(x_n, \theta)||^2 \tag{10}$$

Problem (10) resembles the sieve least-square problem when the number of sigmoidal components R increase appropriately with the same size N. We refer detail discussions on asymptotic normality and convergence properties of sieve least-square estimator to Chapter 3 in Chen (2007).

3 Simulation study

To demonstrate the proposed model and estimator, we design a toy perturbed utility route choice model with $R = 2,\ E = 6$, unit weight $w = \mathbf{1}$, and other parameters specified as follows:

$$\beta^* = (0.5, 1.0),\ \mu^* = (5.0, 2.0),\ \alpha^* = (0.5, 1.0), \gamma^* = (-0.5, 0.5).$$

Besides, $\sigma(\cdot)$ is specified as the logistic function, and each attribute $z_{e,d}$ is uniformly sampled from the unit interval. We then generate choice probabilities by solving x^* from problem (2) and estimate parameters using samples (x^*, z) while fixing $\beta_2 = 1.0$.

Table 1 – *Mean and std. (in brackets) of parameter estimates over 20 replications.*

Sample size	β_1	μ_1	α_1	γ_1	μ_2	α_2	γ_2	**RMSE** $\sqrt{N}$·**RMSE**
50	0.4977 (0.0068)	4.9196 (0.1717)	0.5014 (0.0031)	-0.5179 (0.0420)	2.0331 (0.0983)	0.9959 (0.0103)	0.5003 (0.0015)	0.2213 1.5652
100	0.4955 (0.0069)	4.9790 (0.0899)	0.4996 (0.0006)	-0.4933 (0.0074)	1.9655 (0.0138)	1.0018 (0.0019)	0.4998 (0.0006)	0.1004 1.0038
200	0.5003 (0.0007)	5.0517 (0.0291)	0.4990 (0.0006)	-0.4858 (0.0065)	1.9660 (0.0134)	1.0037 (0.0016)	0.4993 (0.0006)	0.0715 1.0115
1000	0.4992 (0.0011)	5.0045 (0.0229)	0.4997 (0.0005)	-0.4957 (0.0077)	1.9854 (0.0223)	1.0011 (0.0020)	0.4998 (0.0003)	0.0366 1.1566
2000	0.5002 (0.0010)	4.9973 (0.0180)	0.5000 (0.0003)	-0.5005 (0.0035)	2.0008 (0.0100)	0.9999 (0.0009)	0.5000 (0.0002)	0.0211 0.9455

As shown in Table 1, the parameter estimates largely recover the true values. Moreover, the RMSE decreases the sample size. Specifically, when $N > 100$, the values of $\sqrt{N}$·RMSE stabilize around 1, which implies that RMSE decreases at the rate of $1/\sqrt{N}$ and demonstrates the asymptotic normality property as per Corollary 3.2 in Chen & White (1999).

In the full paper, we will further demonstrate the capability of the proposed PUM to represent individual demand using real data and exemplify the estimation efficiency. In addition, we will compare it to other nonparametric and semi-nonparametric methods.

References

Albertini, Francesca, Sontag, Eduardo D, & Maillot, Vincent. 1993. Uniqueness of weights for neural networks. *Artificial neural networks for speech and vision*, **3**.

Allen, Roy, & Rehbeck, John. 2019. Identification with additively separable heterogeneity. *Econometrica*, **87**(3), 1021–1054.

Chen, Xiaohong. 2007. Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, **6**, 5549–5632.

Chen, Xiaohong, & White, Halbert. 1999. Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Transactions on Information Theory*, **45**(2), 682–691.

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

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

Fosgerau, Mogens, Monardo, Julien, & de Palma, André. 2024. The Inverse Product Differentiation Logit Model. *American Economic Journal: Microeconomics.*

Grenander, Ulf. 1981. Abstract inference.

McFadden, Daniel. 1981. Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, **198272**.