Journal of Industrial and
Management Optimization

https://www.aimspress.com/journal/jimo

JIMO, 22(3): 1394–1418.
DOI:10.3934/jimo.2026051
Received: 23 September 2025
Revised: 09 January 2026
Accepted: 22 January 2026
Published: 12 February 2026

Research article

Distributionally robust optimization framework for attribute-independent preference estimation

Lingyun Ji and Dali Zhang*

Antai College of Economic and Management, Shanghai Jiao Tong University, 200030, Shanghai, China

* **Correspondence:** Email: zhangdl@sjtu.edu.cn.

Appendices

A. Proof

Proof of Proposition 2.1

Proof. This proposition establishes the inverse relationship between deterministic utilities and choice probabilities under the multinomial logit (MNL) model. The analysis starts from the classic MNL formula, which assumes the random utility components ε_k are identically distributed (i.i.d.) Gumbel(0,1) distributed [1, 2]:

$$p_k = \frac{e^{V_k}}{\sum_{l\in\mathcal{K}} e^{V_l}}. \tag{A.1}$$

Our goal is to express V_k in terms of p_k.

For the benchmark alternative $k = 0$, the formula is:

$$p_0 = \frac{e^{V_0}}{\sum_{l\in\mathcal{K}} e^{V_l}}. \tag{A.2}$$

By dividing Equation (A.1) by Equation (A.2), the denominator cancels out, yielding:

$$\frac{p_k}{p_0} = \frac{e^{V_k}}{e^{V_0}}.$$

Taking the natural logarithm of both sides gives the relationship between the utility differences and the probability ratios:

$$\ln\left(\frac{p_k}{p_0}\right) = V_k - V_0.$$

By setting the benchmark utility $V_0 = 0$ for normalization, we obtain the first result of the proposition:

$$V_k = \ln\left(\frac{p_k}{p_0}\right). \tag{A.3}$$

For the second part, the cumulative distribution function (CDF) of U_k is computed as follows, using the definition $U_k = V_k + \varepsilon_k$ and the standard Gumbel CDF:

$$\begin{aligned} P(U_k \le u) &= P(V_k + \varepsilon_k \le u) \\ &= P(\varepsilon_k \le u - V_k) \\ &= e^{-e^{-\left(u - \ln\left(\frac{p_k}{p_0}\right)\right)}} \\ &= e^{-\frac{p_k}{p_0} e^{-u}}. \end{aligned}$$

This completes the proof. □

Proof of Proposition 2.2

Proof. The objective of the inner minimization problem is the worst-case expected utility $\mathbb{E}^{\mathbb{P}}\left[\tilde{U}_m\right]$, where the random variable $\tilde{U}_m$ is defined as $\max_{k \in \mathcal{K}_m} U_k$. The set of available alternatives $\mathcal{K}_m$ is a random term determined by the strategy θ^m.

To compute the expectation, we apply the law of total expectation by conditioning on the realization of the random set $\mathcal{K}_m$. Let $\mathcal{K}'$ be a specific non-empty subset of $\mathcal{K}$ (i.e., $\mathcal{K}' \in \mathscr{K}$).

$$\mathbb{E}^{\mathbb{P}}\left[\tilde{U}_m\right] = \sum_{\mathcal{K}' \in \mathscr{K}} P(\mathcal{K}_m = \mathcal{K}') \cdot \mathbb{E}\left[\tilde{U}_m \mid \mathcal{K}_m = \mathcal{K}'\right]. \tag{A.4}$$

The probability of realizing a specific set $\mathcal{K}'$ is given by the independent availability assumption:

$$P(\mathcal{K}_m = \mathcal{K}') = \prod_{k_1 \in \mathcal{K}'} \theta^m_{k_1} \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta^m_{k_2}). \tag{A.5}$$

Next, we evaluate the conditional expectation term. When we condition on the event $\{\mathcal{K}_m = \mathcal{K}'\}$, the random variable $\tilde{U}_m = \max_{k \in \mathcal{K}_m} U_k$ becomes the specific random variable $\max_{k \in \mathcal{K}'} U_k$. Therefore, the conditional expectation is simply the expectation of this new, well-defined random variable:

$$\mathbb{E}\left[\tilde{U}_m \mid \mathcal{K}_m = \mathcal{K}'\right] = \mathbb{E}\left[\max_{k \in \mathcal{K}'} U_k\right]. \tag{A.6}$$

Note that the value of this expectation depends on set $\mathcal{K}'$.

A foundational result in discrete choice theory states that for a set of random variables $U_k = V_k + \varepsilon_k$, where the ε_k is i.i.d. standard Gumbel, the expected value of their maximum is given by the celebrated 'log-sum formula' [1, 3]:

$$\mathbb{E}\left[\max_{k \in \mathcal{K}'} U_k\right] = \ln\left(\sum_{k \in \mathcal{K}'} e^{V_k}\right) + \gamma, \tag{A.7}$$

where γ is the Euler–Mascheroni constant.

From Proposition 2.1, we have the relationship $V_k = \ln(p_k/p_0)$, which implies $e^{V_k} = p_k/p_0$. Substituting this into the log-sum formula (A.7) yields

$$\begin{aligned}\mathbb{E}\left[\max_{k\in\mathcal{K}'} U_k\right] &= \ln\left(\sum_{k\in\mathcal{K}'}\frac{p_k}{p_0}\right)+\gamma \\ &= \ln\left(\frac{\sum_{k\in\mathcal{K}'} p_k}{p_0}\right)+\gamma.\end{aligned} \tag{A.8}$$

Finally, substituting the results from (A.5) and (A.8) back into (A.4), we obtain the objective function as stated in the proposition. This completes the proof. □

Proof of Proposition 3.1

Proof. To establish the convexity of $f(t_0, t_1, \ldots, t_K)$, we first present the following lemma:

Lemma A.1. *A symmetric matrix $H \in \mathbb{R}^{N\times N}$ with elements H_{ij} for $i, j = 1, 2, \ldots, N$ is positive semi-definite if the following conditions hold:*

- $H_{ii} \geq 0$ *for all* i,
- $H_{ij} \leq 0$ *for* $i \neq j$,
- $\sum_{j=1}^{N} H_{ij} = 0$ *for all* $i = 1, \ldots, N$.

Proof of Lemma A.1

Proof. Consider $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$,

$$\begin{aligned} x^T H x &= \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j H_{ij} \\ &= \sum_{i=1}^{N} x_i^2 H_{ii} + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} x_i x_j H_{ij} \\ &= \sum_{i=1}^{N} x_i^2\left(-\sum_{j=i+1}^{N} H_{ij}\right) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} x_i x_j H_{ij} \\ &= -\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} H_{ij}(x_i - x_j)^2. \end{aligned} \tag{A.9}$$

Note that with $H_{ij} \leq 0$ and $(x_i - x_j)^2 \geq 0$ for $i \neq j$, then $x^T H x \geq 0$, confirming that H is positive semi-definite. □

Now for the convexity of function f, define $\mathscr{K}(i) = \{\mathcal{K}' \mid \mathcal{K}' \in \mathscr{K},\ i \in \mathcal{K}'\}$, $\mathscr{K}'(i) = \{\mathcal{K}' \mid \mathcal{K}' \in \mathscr{K},\ i \notin \mathcal{K}'\}$. The first-order derivative of f with respect to t_i, where $i \in \mathcal{K}$, is given by

$$\begin{aligned}\frac{\partial f}{\partial t_i} &= \frac{\partial}{\partial t_i} \sum_{\mathcal{K}' \in \mathscr{K}} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \left(\gamma + \ln \sum_{k \in \mathcal{K}'} e^{t_k} - \ln e^{t_0} \right) \\ &= \frac{\partial}{\partial t_i} \sum_{\mathcal{K}' \in \mathscr{K}(i)} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \left(\gamma + \ln \sum_{k \in \mathcal{K}'} e^{t_k} - \ln e^{t_0} \right) \\ &+ \frac{\partial}{\partial t_i} \sum_{\mathcal{K}' \in \mathscr{K}'(i)} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \left(\gamma + \ln \sum_{k \in \mathcal{K}'} e^{t_k} - \ln e^{t_0} \right) \\ &= \sum_{\mathcal{K}' \in \mathscr{K}(i)} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \left(\frac{e^{t_i}}{\sum_{k \in \mathcal{K}'} e^{t_k}} - 1_{\{i=0\}} \right),\end{aligned} \tag{A.10}$$

where the second term on the right-hand side of the second equation in (A.10) is zero, and $1_{\{i=0\}}$ is 1 when $i = 0$ and 0 otherwise. Next, we consider the second-order derivative:

$$\begin{aligned}\frac{\partial^2 f}{\partial {t_i}^2} &= \prod_{k_1 \in \{i\}} \theta_{k_1}^m \, q\,(\mathcal{K}/\{i\},\ \theta^m)\, \frac{e^{t_i} \sum_{l \in \{i\}} e^{t_l} - e^{2t_i}}{(\sum_{k \in \mathcal{K}'} e^{t_k})^2} \\ &+ \sum_{\substack{\mathcal{K}' \in \mathscr{K}(i) \\ \mathcal{K}' \neq \{i\}}} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \frac{e^{t_i} \sum_{l \in \mathcal{K}'} e^{t_l} - e^{2t_i}}{(\sum_{k \in \mathcal{K}'} e^{t_k})^2} \\ &= \sum_{\substack{\mathcal{K}' \in \mathscr{K}(i) \\ \mathcal{K}' \neq \{i\}}} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \frac{\sum_{l \in \mathcal{K}'/\{i\}} e^{t_i + t_l}}{(\sum_{k \in \mathcal{K}'} e^{t_k})^2} \\ &= \sum_{\substack{\mathcal{K}' \in \mathscr{K}(i) \\ \mathcal{K}' \neq \{i\}}} \sum_{l \in \mathcal{K}'/\{i\}} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \frac{e^{t_i + t_l}}{(\sum_{k \in \mathcal{K}'} e^{t_k})^2}.\end{aligned} \tag{A.11}$$

Recall that for $i, l \in \mathcal{K}$, we have $\mathscr{K}(i) = \{\mathcal{K}' \mid \mathcal{K}' \in \mathscr{K}, i \in \mathcal{K}'\}$ and $\mathscr{K}(l) = \{\mathcal{K}' \mid \mathcal{K}' \in \mathscr{K}, l \in \mathcal{K}'\}$. Then the $\mathcal{K}'$ that determines the coefficients of $e^{t_i + t_l}$ in (A.11) satisfies the following condition:

$$\{\mathcal{K}' \mid \mathcal{K}' \in \mathscr{K}, i \in \mathcal{K}', l \in \mathcal{K}'\} = \mathscr{K}(i) \cap \mathscr{K}(l),$$

and this holds for all $l \in \mathcal{K}$. Then, (A.11) can be regarded as a summation of all the terms $e^{t_i + t_l}$ with corresponding coefficients, for all $l \in \mathcal{K}$,

$$\frac{\partial^2 f}{\partial {t_i}^2} = \sum_{l \in \mathcal{K}/\{i\}} \sum_{\mathcal{K}' \in \mathscr{K}(i) \cap \mathscr{K}(l)} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \frac{e^{t_i + t_l}}{(\sum_{k \in \mathcal{K}'} e^{t_k})^2}. \tag{A.12}$$

For the second-order partial derivatives,

$$\frac{\partial^2 f}{\partial t_i \partial t_j} = - \sum_{\mathcal{K}' \in \mathscr{K}(i) \cap \mathscr{K}(j)} \prod_{k_1 \in \mathcal{K}'} \theta_{k_1}^m \prod_{k_2 \in \mathcal{K} \setminus \mathcal{K}'} (1 - \theta_{k_2}^m) \frac{e^{t_i + t_j}}{(\sum_{k \in \mathcal{K}'} e^{t_k})^2}\,, \quad i \neq j. \tag{A.13}$$

Therefore,

$$\frac{\partial^2 f}{\partial {t_i}^2} + \sum_{j\in\mathcal{K}/\{i\}} \frac{\partial^2 f}{\partial t_i \partial t_j} = 0,\ \forall i \in \mathcal{K}.$$

Note that $\theta_i \in [0,1]$ for each $i \in \mathcal{K}$, $\prod_{i\in\mathcal{K}} \theta_i > 0$, and $\prod_{i\in\mathcal{K}}(1-\theta_i) > 0$. Therefore, the second-order partial derivatives satisfy $\frac{\partial^2 f}{\partial t_i \partial t_j} \le 0$ and $\frac{\partial^2 f}{\partial t_i^2} \ge 0$. The Hessian matrix H of function f has the following characteristics:

i) $H_{ij} = H_{ji}$ for all $i, j \in \mathcal{K}$.

ii) $H_{ii} \ge 0$ and $H_{ij} \le 0$ for all $i, j \in \mathcal{K}$, where $i \neq j$.

iii) $\sum_{j\in\mathcal{K}} H_{ij} = 0$ for all $i \in \mathcal{K}$.

Thus, H is a positive semi-definite matrix as presented in Lemma A.1, implying that $f(t_0, t_1, \ldots, t_K)$ is a convex function. □

Proof of Lemma 3.1

Proof. Consider $i \in \{0, \ldots, K\}$,

$$\frac{\partial g_i}{\partial t_i} = \frac{e^{t_i} \sum_{l=0}^K e^{t_l} - e^{2t_i}}{\left(\sum_{l=0}^K e^{t_l}\right)^2} = \frac{e^{t_i}(\sum_{l=0}^K e^{t_l} - e^{t_i})}{\left(\sum_{l=0}^K e^{t_l}\right)^2}. \tag{A.14}$$

For the second-order derivative, we have

$$\begin{aligned} \frac{\partial^2 g_i}{\partial t_i^2} &= \frac{e^{t_i}(\sum_{l=0}^K e^{t_l} - e^{t_i})\left(\sum_{l=0}^K e^{t_l}\right)^2}{\left(\sum_{l=0}^K e^{t_l}\right)^4} - \frac{e^{t_i}(\sum_{l=0}^K e^{t_l} - e^{t_i}) * 2(\sum_{l=0}^K e^{t_l}) * e^{t_i}}{\left(\sum_{l=0}^K e^{t_l}\right)^4} \\ &= \frac{e^{t_i}(\sum_{l=0}^K e^{t_l} - e^{t_i}) \sum_{l=0}^K e^{t_l} - e^{t_i}(\sum_{l=0}^K e^{t_l} - e^{t_i}) * 2e^{t_i}}{\left(\sum_{l=0}^K e^{t_l}\right)^3} \\ &= \frac{e^{t_i}(\sum_{l=0}^K e^{t_l} - 2e^{t_i})}{\left(\sum_{l=0}^K e^{t_l}\right)^3} \left[\sum_{l=0}^K e^{t_l} - e^{t_i} \right]. \end{aligned} \tag{A.15}$$

For the second-order partial derivatives,

$$\begin{aligned} \frac{\partial^2 g_i}{\partial t_i \partial t_j} &= \frac{e^{t_i+t_j}\left(\sum_{l=0}^K e^{t_l}\right)^2}{\left(\sum_{l=0}^K e^{t_l}\right)^4} - \frac{e^{t_i}(\sum_{l=0}^K e^{t_l} - e^{t_i}) * 2(\sum_{l=0}^K e^{t_l}) * e^{t_j}}{\left(\sum_{l=0}^K e^{t_l}\right)^4} \\ &= \frac{e^{t_i} \sum_{l=0}^K e^{t_l} - e^{t_i}(\sum_{l=1}^N 2e^{t_l} - 2e^{t_i})}{\left(\sum_{l=0}^K e^{t_l}\right)^3} e^{t_j} \\ &= -\frac{e^{t_i}(\sum_{l=0}^K e^{t_l} - 2e^{t_i})}{\left(\sum_{l=0}^K e^{t_l}\right)^3} e^{t_j}, \quad j \neq i\ . \end{aligned} \tag{A.16}$$

Therefore,

$$
\begin{aligned}
\frac{\partial^2 g_i}{\partial t_i^2} + \sum_{\substack{j=0\\ j\neq i}}^{K} \frac{\partial^2 g_i}{\partial t_i \partial t_j} &= \frac{e^{t_i}(\sum_{l=0}^{K} e^{t_l} - 2e^{t_i})}{(\sum_{l=0}^{K} e^{t_l})^3} \left[\sum_{l=0}^{K} e^{t_l} - e^{t_i}\right] - \sum_{\substack{j=0\\ j\neq i}}^{K} \frac{e^{t_i}(\sum_{l=0}^{K} e^{t_l} - 2e^{t_i})}{(\sum_{l=0}^{K} e^{t_l})^3} e^{t_j} \\
&= \frac{e^{t_i}(\sum_{l=0}^{K} e^{t_l} - 2e^{t_i})}{(\sum_{l=0}^{K} e^{t_l})^3} \left[\sum_{l=0}^{K} e^{t_l} - \sum_{i=0}^{K} e^{t_i}\right] \\
&= 0.
\end{aligned}
$$

It is straightforward that

$$
\sum_{l=0}^{K} e^{t_l} - e^{t_i} = \sum_{\substack{j=0\\ j\neq i}}^{K} e^{t_j} > 0,
$$

which shows that

$$
\frac{\partial^2 g_i}{\partial t_i \partial t_j} < 0, \ \frac{\partial^2 g_i}{\partial t_i^2} > 0
$$

is equivalent to

$$
\sum_{l=0}^{K} e^{t_l} - 2e^{t_i} > 0.
$$

Therefore, by Lemma A.1, the function $g_i(t)$ is convex in $\mathbb{R}^{K+1}$ if its corresponding probability term satisfies:

$$
p_i = \frac{e^{t_i}}{\sum_{l=0}^{K} e^{t_l}} < \frac{1}{2}, \quad \forall i \in \{0, \ldots, K\}.
$$

□

Proof of Lemma 3.2

Proof. First,

$$\psi'(x) = \log\left(\frac{x}{a}\right) + x \cdot \frac{1}{x} = \log\left(\frac{x}{a}\right) + 1.$$

By solving $\psi'(x) = 0$, we find that the solution is $x^* = \frac{a}{e}$. When $x \in \left(0, \frac{a}{e}\right)$, $\psi'(x) < 0$. When $x \in \left(\frac{a}{e}, +\infty\right)$, $\psi'(x) > 0$. Therefore, $\psi(x) = x \ln \frac{x}{a}$, where $x > 0$ and $a \in (0, 1)$, is decreasing in $\left(0, \frac{a}{e}\right)$ and increasing in $\left(\frac{a}{e}, +\infty\right)$. Moreover,

$$\psi''(x) = \frac{1}{x} > 0,$$

which demonstrates that $\psi(x)$ is a convex function.

□

Proof of Proposition 3.2

Proof. We first rewrite the function $c(t_0, t_1, \ldots, t_K)$ as

$$c(t_0, t_1, \ldots, t_K) = \sum_{i \in \mathcal{K}} h_i \left(g_0(t), g_1(t), \ldots, g_K(t)\right) - \eta, \tag{A.17}$$

where

$$h_i \left(g_0(t), g_1(t), \ldots, g_K(t)\right) := \sum_{k \in \mathcal{K}} g_k(t) \ln \frac{g_k(t)}{\hat{p}_k} - \eta, \quad g_i(t) := \frac{e^{t_i}}{\sum_{l \in \mathcal{K}} e^{t_l}},$$

for $t = (t_0, t_1, \ldots, t_K)$.

Note that from Lemma 3.1, g_k is convex under the condition that $\frac{e^{t_k}}{\sum_{l \in \mathcal{K}} e^{t_l}} < \frac{1}{2}$ for all $k \in \mathcal{K}$ in (3.6). Moreover, due to $\frac{e^{t_i}}{\sum_{l \in \mathcal{K}} e^{t_l}} > \frac{\hat{p}_i}{e}$ in (3.6), the monotonicity and convexity of h_i are ensured based on Lemma 3.2. According to Boyd and Vandenberghe [4], the function $c(t_0, t_1, \ldots, t_K)$ is convex if and only if, for all h_i and g_k, where $i, k \in \mathcal{K}$, both are convex and h_i is nondecreasing in each argument.

□

B. Algorithm

Algorithm 1: Augmented Lagrangian Method

Initialize the multiplier μ^0, the penalty factor $\sigma_0 > 0$, the constraint violation threshold $\varepsilon > 0$, the accuracy tolerance $\delta > 0$, and constants $0 < \alpha \leq \beta \leq 1$ and $\rho > 1$. Set $\delta_0 = \frac{1}{\sigma_0}$, $\varepsilon_0 = \frac{1}{\sigma_0^{\alpha}}$, and $j = 0$.

Given $\hat{p}$ in 2.2, generate $p^0 \in \mathbb{R}^{K+1}$, such that:

$$p_k^0 \in [0,1],\ k \in \mathcal{K},\ \sum_{k\in\mathcal{K}} p_k^0 \ln \frac{p_k^0}{\hat{p}_k} \leq \eta.$$

Transform p^0 into $t^0 \in \mathbb{R}^{K+1}$ defined by $t_k^0 = \ln p_k^0$ for $k \in \mathcal{K}$.

for $j = 0, 1, 2, \ldots$ **do**

Solve the following subproblem:

$$\min_{t\in\mathbb{R}^{K+1}} \check{L}_{\sigma_j}(t, \mu^j) \tag{B.1}$$

starting from t^j (using t^0 for the first iteration $j = 0$), and stop when the solution t^{j+1} satisfies:

$$\|\nabla_t \check{L}_{\sigma_j}(t^{j+1}, \mu^j)\|_2 \leq \delta_j. \tag{B.2}$$

if $v_j(t^{j+1}) \leq \varepsilon_j$ **then**

if $v_j(t^{j+1}) \leq \varepsilon$ ***and*** $\|\nabla_t \check{L}_{\sigma_j}(t^{j+1}, \mu^j)\|_2 \leq \delta$ **then**

Obtain an approaching solution (t^{j+1}, μ^j) and **stop**.

end

Update multipliers:

$$\begin{aligned} \mu_s^{j+1} &= \max\{\mu_s^j + \sigma_j c_s(t^{j+1}), 0\}, \\ \mu_{r_k}^{j+1} &= \max\{\mu_{r_k}^j + \sigma_j c_{r_k}(t^{j+1}), 0\}, \\ \mu_{o_k}^{j+1} &= \max\{\mu_{o_k}^j + \sigma_j c_{o_k}(t^{j+1}), 0\}. \end{aligned} \tag{B.3}$$

Keep the penalty factor unchanged: $\sigma_{j+1} = \sigma_j$.

Update the solution error and constraint violation for the subproblem (B.1):

$$\delta_{j+1} = \frac{\delta_j}{\sigma_{j+1}},\ \varepsilon_{j+1} = \frac{\varepsilon_j}{\sigma_{j+1}^{\beta}}.$$

else

Keep the multipliers unchanged: $\mu^{j+1} = \mu^j$, and update the penalty factor: $\sigma_{j+1} = \rho\sigma_j$.

Update the solution error and constraint violation for the subproblem (B.1):

$$\delta_{j+1} = \frac{1}{\sigma_{j+1}},\ \varepsilon_{j+1} = \frac{\varepsilon_j}{\sigma_{j+1}^{\alpha}}.$$

end

end

In the algorithm, we first find the optimal solution t^{j+1} of subproblem 3.17 and then use $v_j(t^{j+1})$ to evaluate the constraint violation of the solution at each iteration. If the specified accuracy level ε_j is achieved, the algorithm updates the multiplier μ^j to μ^{j+1}, improving the accuracy of the subproblem, while the penalty factor σ_j remains unchanged. On the contrary, if the given accuracy level ε_j is not achieved, the algorithm keeps the multiplier unchanged and increases the penalty factor appropriately to find a solution with a smaller constraint violation in the next iteration.

C. Project Investment Problem

In a competitive automotive market, a manufacturer strategically allocates limited capital resources across several innovative projects. These projects are designed to improve various attributes of the manufacturer's automotive products, potentially leading to changes in pricing (see Section 6.2 in Hu et al. [5]). For instance, the Noise, Vibration, and Harshness (NVH) project impacts attributes, including comfort, depreciation rate, and price, while the engine upgrade project affects fuel consumption, acceleration, and dealership coverage (number of dealers), among others. By improving key attributes, such as safety, comfort, depreciation rate, fuel consumption, acceleration, dealership coverage, and so forth, the manufacturer aims to meet dynamic demands, increase market share, and enhance brand influence. However, the outcomes of the investments in these projects are uncertain, and the precise impact of each project on various attributes is difficult to quantify. Therefore, we employ the sequential decision framework proposed in our study to navigate these uncertainties.

We present nine investment projects, along with an additional no-investment option, as shown in Table 1. Since the numerical data provided in Hu et al. [5] do not match our utility formulation based on the choice probabilities, we construct simulation data using the background and operational logic in their study. Each project has a corresponding historical selection probability. We also provide ten investment portfolios in Table 2, each designed to allocate capital differently among the projects, resulting in varying success rates. For each investment project, the success rate is influenced by the amount of capital allocated. Generally, as the capital increases, the success rate also rises. The success rate for each portfolio is estimated by the Research and Development department, although the detailed estimation procedure is beyond the scope of this study. Due to the overall capital limit, only a subset of these projects is selected in each investment portfolio, with one project chosen for further commitment in the second stage after resolving market uncertainties.

Table 1. The set of investment project alternatives and their simulated selection probabilities ($\hat{p}$), constructed using the background described in Hu et al. [5] for the Project Investment Problem case study.

No.	Investment Project	Description	Selection Prob. ($\hat{p}$)
0	**None of These**	No-investment option.	4%
1	**Safety Promotion**	Design a new structure to alleviate the impacts of the collision.	4%
2	**New Car Model Development**	Create a concept car with the aim of a more fashion style and higher market acceptance before actually producing it.	7%
3	**Engine Upgrade**	Upgrade the current engine and its vibration sensor system to achieve more horsepower and less engine noise.	7%
4	**E-Platform Development**	Improve the human-vehicle interaction experience and visualize the performances of the car.	11%
5	**Computational Fluid Dynamics Testing System Development**	Implement related fluid mechanics and numerical analysis into a testing platform to analyze the performance of concept cars.	10%
6	**Common Modular Platform Development**	Develop a platform used for subcompact and compact car models with internal combustion engine and battery-electric cars.	12%
7	**Noise, Vibration, Harshness Digitalization**	Incorporate the characteristics of the noise, vibration and harshness of vehicles to a digital platform with the target of more efficient computation in evaluating the driver satisfactions.	28%
8	**Driving Assistance System Development**	Incorporate the latest interface standards and running multiple vision-based algorithms to support real-time multimedia, vision co-processing, and sensor fusion subsystems.	3%
9	**Digitalization of Marketing Network**	Build a digital platform used for all the dealers across different regions to share the customer resources, the service standards and the marketing strategies.	11%

The first column in Table 2 represents the investment project numbers, ranging from 0 to 9. The subsequent columns display the success rates of the corresponding portfolios, as indicated by the numbers in the headers beneath *Portfolio*. Project 0, which corresponds to the no-investment option that maintains the original setting, has a success rate of 100%. This success rate remains consistent across all portfolios, as it is unaffected by capital allocation. A success rate of 0% in Table 2 indicates that the portfolio in that column allocates no capital to the project in that row. Additionally, success rates between 0% and 100% indicate that some capital is allocated to the project, although the outcome remains uncertain.

Table 2. The success probabilities θ_k^m for each investment project k under each of the ten investment portfolios m in the Project Investment Problem case study, computed using simulated data.

Investment	**Portfolio**									
Project No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
1	72%	0%	75%	0%	74%	80%	78%	0%	70%	0%
2	0%	64%	0%	65%	0%	67%	0%	69%	0%	82%
3	0%	79%	60%	0%	0%	0%	75%	81%	0%	0%
4	60%	0%	62%	0%	71%	85%	0%	0%	63%	0%
5	0%	66%	0%	70%	0%	0%	68%	0%	0%	67%
6	35%	0%	72%	73%	0%	0%	0%	71%	74%	0%
7	80%	40%	0%	81%	0%	0%	83%	0%	0%	84%
8	0%	61%	0%	0%	70%	0%	0%	73%	67%	0%
9	0%	0%	0%	0%	62%	74%	0%	63%	0%	65%

The optimal solutions are obtained by solving problem (3.7) with the corresponding parameters for various portfolios in Table 2. The values in Row *Utility* in Table 3 represent the expected worst-case utilities that the manufacturer can achieve in the second stage, based on the success rates of the investment projects. From these values, we observe that Portfolio 4 yields the maximum worst-case utility and is therefore selected in the first stage. The optimal values for these ten portfolios range from 0.750 to 1.374. It is important to note that these results are based on the assumption that the deterministic part of the utility for the no-investment option is 0, i.e., $V_0 = 0$. This implies that these values serve only for comparison when selecting the best investment portfolio and lack practical significance.

Table 3. The optimal worst-case expected utility for each of the ten investment portfolios with an ambiguity radius of $\eta = 0.1$, computed using simulated data.

Portfolio	1	2	3	4	5	6	7	8	9	10
Utility	1.199	1.057	0.855	1.374	0.750	0.924	1.270	1.039	0.783	1.371

Figure 1 illustrates the convergence of the algorithm across portfolios. The bold red line represents the convergence of the algorithm for the optimal portfolio (Portfolio 4), while the other lines represent the convergence of the algorithm for the other portfolios.

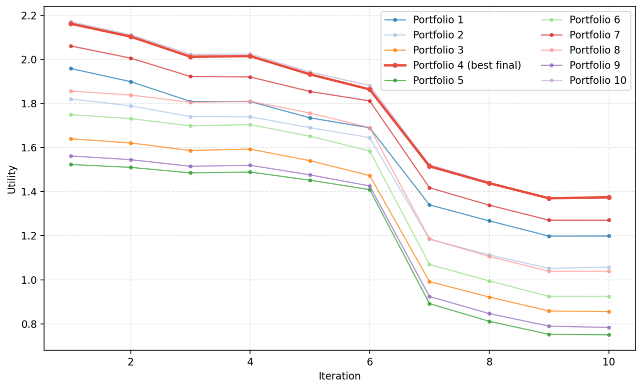

Figure 1. Convergence of the worst-case expected utility for the ten investment portfolios with an ambiguity radius of $\eta = 0.1$, computed using simulated data.

Since it uses the same hyperparameters as in Section 4.1, it exhibits similar convergence behavior to that shown in Figure 2, although the number of iterations required for convergence may vary across datasets.Both experiments converge at an efficient rate, as the differences between each iterate are significantly reduced within only a few iterations.

D. Synthetic Data

In this section, we provide the details of the synthetic data used in Section 4.4. For each value of $K \in \{4, 5, 6, 7, 8, 9, 10\}$, we consider a single strategy and provide the choice probabilities ($\hat{p}$) and availability probabilities (θ^1) for each K, as follows.

Alternative $K = 4$ The choice probabilities ($\hat{p}$) are:

$$\hat{p} = [0.01, 0.48, 0.30, 0.21]. \tag{D.1}$$

The availability probabilities (θ^1) are:

$$\theta^1 = [100\%, 30\%, 70\%, 0\%]. \tag{D.2}$$

Alternative $K = 5$ The choice probabilities ($\hat{p}$) are:

$$\hat{p} = [0.05, 0.23, 0.35, 0.20, 0.17]. \tag{D.3}$$

The availability probabilities (θ^1) are:

$$\theta^1 = [100\%, 80\%, 33\%, 0\%, 40\%]. \tag{D.4}$$

Alternative $K = 6$ The choice probabilities ($\hat{p}$) are:

$$\hat{p} = [0.41, 0.32, 0.01, 0.09, 0.14, 0.03]. \tag{D.5}$$

The availability probabilities (θ^1) are:

$$\theta^1 = [100\%, 70\%, 60\%, 50\%, 0\%, 10\%]. \tag{D.6}$$

Alternative $K = 7$ The choice probabilities ($\hat{p}$) are:

$$\hat{p} = [0.19, 0.22, 0.32, 0.01, 0.09, 0.14, 0.03]. \tag{D.7}$$

The availability probabilities (θ^1) are:

$$\theta^1 = [100\%, 0\%, 70\%, 30\%, 50\%, 10\%, 0\%]. \tag{D.8}$$

Alternative $K = 8$ The choice probabilities ($\hat{p}$) are:

$$\hat{p} = [0.19, 0.22, 0.12, 0.20, 0.01, 0.09, 0.14, 0.03]. \tag{D.9}$$

The availability probabilities (θ^1) are:

$$\theta^1 = [100\%, 90\%, 60\%, 0\%, 20\%, 0\%, 10\%, 0\%]. \tag{D.10}$$

Alternative $K = 9$ The choice probabilities ($\hat{p}$) are:

$$\hat{p} = [0.02, 0.12, 0.15, 0.20, 0.10, 0.08, 0.11, 0.17, 0.05]. \tag{D.11}$$

The availability probabilities (θ^1) are:

$$\theta^1 = [100\%, 85\%, 75\%, 90\%, 65\%, 0\%, 0\%, 0\%, 0\%]. \tag{D.12}$$

Alternative $K = 10$ The choice probabilities ($\hat{p}$) are:

$$\hat{p} = [0.04, 0.04, 0.07, 0.07, 0.10, 0.10, 0.12, 0.28, 0.03, 0.11]. \tag{D.13}$$

The availability probabilities (θ^1) are:

$$\theta^1 = [100\%, 72\%, 0\%, 0\%, 60\%, 0\%, 35\%, 80\%, 0\%, 0\%]. \tag{D.14}$$

References

1. D. McFadden, Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics.*, Academic Press, New York, 1974, 105–142. https://escholarship.org/uc/item/61s3q2xr
2. H. Theil, A multinomial extension of the linear logit model, *Advanced Studies in Theoretical and Applied Econometrics,* Springer, Dordrecht, 1992, 181–191. https://doi.org/10.1007/978-94-011-2546-8_11
3. M. Ben-Akiva, S. R. Lerman, *Discrete choice analysis*, MIT Press, USA, 1985.
4. S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004.
5. J. Hu, D. L. Zhang, H. F. Xu, S. N. Zhang, Distributional utility preference robust optimization models in multi-attribute decision making, *Math. Program.*, **212** (2025), 519–565. https://doi.org/10.1007/s10107-024-02114-y