Omega 142 (2026) 103514

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

An exact scenario-independent deterministic equivalent form of stochastic programs embedding Multivariate Extreme Value discrete choice problems ☆

Michel Bierlaire [a], Edoardo Fadda [b], Lohic Fotio Tiotsop [c], Daniele Manerba [d],*

[a] Transport and Mobility Lab., Ecole Polytechnique Fédérale de Lausanne, CH-1015, Lausanne , Switzerland
[b] Department of Mathematical Sciences, Politecnico di Torino, corso Duca degli Abruzzi 24, Torino, Italy
[c] Department of Control and Computer Engineering, Politecnico di Torino, corso Duca degli Abruzzi 24, Torino, Italy
[d] Department of Information Engineering, Università degli Studi di Brescia, via Branze 38, Brescia, Italy

ARTICLE INFO

Keywords:
Stochastic programming
Discrete choice problems
Multivariate extreme value models
Deterministic equivalent form
Logit
Nested logit

ABSTRACT

We address the class of two-stage Stochastic Programs embedding, in their second stage, a set of Discrete Choice Problems (tsSP-DCPs), one independent from the other, but all linked by the first-stage decisions This decisional structure can be found within many managerial and organizational contexts in relation to several applications such as location–allocation, routing, scheduling, and sequencing. Generally, solving a two-stage stochastic program requires the analytical derivation of the second-stage problem's expected optimum, which in turn implies calculating a multidimensional integral. Therefore, a common practice is approximating the random variables involved through a finite set of scenarios and solving a huge scenario-dependent program, which affects the scalability of making optimal decisions under uncertainty. However, under some assumptions commonly adopted in the discrete choice context, we can prove that a closed-form analytical expression of the expected second-stage optimum of a tsSP-DCP can be derived, and an exact scenario-independent equivalent deterministic program can be obtained. Through a numerical showcase, we validate our approach in terms of efficiency and effectiveness. Our equivalent deterministic form, which only requires estimating a few parameters in practice, is far less computationally demanding than any scenario-based deterministic equivalent forms, thereby simplifying the decision-making process. Finally, we show that our methodology can be generalized to address a larger class of two-stage stochastic programs, i.e., those in which the second-stage expected optimum is decomposable into a finite number of expectations of Extreme Values and in which second-stage utilities may also depend on first-stage decisions.

1. Introduction

Stochastic Programming (SP) is one of the leading modeling paradigms to tackle optimization problems in the presence of uncertain parameters [1,2]. In fact, SP models can well capture the behavior of many decision-making problems by implementing optimization actions that mix both robustness and flexibility features against unknown behaviors of the involved information. The application of such models is nowadays widely used in several organizational fields, with countless examples in finance, logistics, production, and any management in general (see, e.g., [3–5]).

In this paper, we address the so-called *two-stage* SP in which the main decisions arise at two different levels (e.g., long-term tactical decisions and day-by-day operational decisions) and are split into two logical sets: *(i) here-and-now* decisions that correspond to the first-stage variables and must be taken before the uncertainty is resolved; *(ii) recourse* decisions that correspond to the second-stage variables and can be postponed in the future after the uncertain parameters become known. Hence, first-stage decisions are those that are actually made under uncertainty and that cannot be modified afterward. In fact, the optimality and the feasibility of the second-stage decisions depend, apart from the uncertainty realization, also on the first-stage decisions. In the first stage, the decision-maker takes into account the problem uncertainty by incorporating the expected value of the future outcomes (i.e., the second-stage objective value) into the objective function.

Unfortunately, SP models are notoriously very challenging to solve, even when the involved functions show nice properties such as convexity or linearity. Being able to derive an analytical expression of the expected optimum of the second-stage problem is, in general, denied by the complexity of calculating the multidimensional integral from

☆ Area: Data-Driven Optimization This manuscript was processed by Associate Editor Yuhong He.

* Corresponding author.
E-mail address: daniele.manerba@unibs.it (D. Manerba).

https://doi.org/10.1016/j.omega.2026.103514
Received 1 April 2025; Accepted 11 January 2026
Available online 12 January 2026

which it derives and of embedding this evaluation within a combinatorial optimization problem [6]. For this reason, a typical approach for SP problems is to represent (most often, approximate) the future uncertainty through a finite set of scenarios and, in turn, to create a so-called *scenario-based deterministic equivalent problem* embedding scenario-dependent variables and constraints [7]. Since the size and complexity of this deterministic program depend on the number of scenarios considered, it generally becomes computationally intractable as the number of scenarios grows. Moreover, when uncertain parameters are defined through their probability distributions, the scenario generation (or selection) process itself becomes a challenging and critical task.

In this work, we are interested in studying two-stage SP problems in which the second stage consists of a series of many Discrete Choice Problems (DCPs). A DCP is a stochastic optimization problem in which the decision-maker needs to select the most profitable single choice among a finite set of alternatives, each characterized by a stochastic utility. DCPs have been used in different fields, such as marketing and transportation [8,9]. In particular, they are useful to model the behavior of rational individuals in microeconomic settings and their dependency on the possible idiosyncrasies of the decision-maker in terms of the evaluation of the utility of each alternative. Mixing classical two-stage SP problems and DCPs leads to a family of optimization settings with a special structure that we name *two-stage Stochastic Programming embedding Discrete Choice Problems* (tsSP-DCP). In a tsSP-DCP, the set of DCPs to be solved in the second stage depends on the decisions taken in the first stage in a very precise way, namely, if the decision-maker acts on a first-stage decision variable, setting its value to something different from 0, the relative second-stage DCP is triggered. Once triggered, the optimization problems to be solved at the second stage consist of selecting the best alternatives (each characterized by a stochastic utility) to maximize the total expected utility. Note that such an expected utility may proportionally depend on the value of the related first-stage decision variable. Typical applications of tsSP-DCPs can be found, e.g., in facility location–allocation problems, bin-packing problems, multi-modal routing, and scheduling (see, e.g., [10–12]). Moreover, their simple combinatorial structure (basically involving the presence of a *multi-arm bandit* type constraint in the second stage) often occurs as a subproblem in more complex settings. As far as we know, tsSP-DCPs have never been formally studied and approached as a unified class of problems. Therefore, our goal is to exploit the mathematical characterization of this class and show how it can be efficiently addressed through analytical approaches, as previously done for simple optimization settings such as the newsvendor problem. In particular, we exploit the special structure of tsSP-DCPs to derive a closed-form analytical expression for the second-stage value and, in turn, a deterministic program that is equivalent to the stochastic problem without resorting to any kind of scenario representation. Our deterministic equivalent form derivation simply assumes, as commonly done in Discrete Choice Theory, that the choice behavior of the decision-maker is captured by a so-called *Multivariate Extreme Value* (MEV) model [13]. In fact, depending on the specific assumptions made on the stochastic oscillation of the utilities, various discrete choice models can be derived to describe the decision-maker's choice behavior when facing a DCP. Some models allow more flexibility than others, e.g., in terms of correlation between alternatives. The models we consider in this work, the MEVs, are arguably the most general ones and embrace a very large quantity of well-known and widely applied choice models, such as the *logit*, the *nested logit*, and the *cross-nested logit* ones. Indeed, it can be shown that any choice model can be approximated arbitrarily well by an MEV, namely by a cross-nested logit model [14].

The contribution of this work is threefold. First, we propose the theoretical background for approaching the class of stochastic problems called tsSP-DCPs, which embed a sequence of independent DCPs within the second stage of a two-stage SP problem. The tsSP-DCP framework can model several basic decisional problems under uncertainty (e.g., in location, assignment, scheduling, sequencing), but, more importantly, its simple combinatorial structure appears as a subproblem in more complex and realistic situations. Also, note that the tsSP-DCP framework provides a structured way for embedding and linking discrete choice modeling within mathematical programming approaches. Second, under the common assumption of capturing the choice behavior of the decision-maker through an MEV model, we derive a closed-form analytical expression for the expected second-stage value of a tsSP-DCP and formulate a deterministic and scenario-independent equivalent model able to solve the complete problem. The effectiveness and the efficiency of such an approach are validated through extensive computational experiments conducted over a showcase facility location–allocation problem. Third, we show that the approach can be applied to a larger structured class of problems as well as used as an approximation in many practical applications by relaxing some assumptions on the data knowledge, thus largely facilitating the process of decision-making under uncertainty.

The rest of the paper is structured as follows. In Section 2, we review the main literature related to the problems studied in this paper, which are formally defined and exemplified in Section 3. After the presentation of the necessary assumptions on the probability distribution of the uncertain parameters, we derive our MEV-based deterministic equivalent form of a tsSP-DCP in Section 4 and discuss its practical applicability. Section 5 is devoted to validating our method and assessing its effectiveness and efficiency through computational experiments on a large number of instances. Section 6 provides a discussion on possible generalizations of our approach to a wider class of stochastic problems. Finally, conclusions are drawn in Section 7.

2. Literature review

Among the several paradigms existing to deal with uncertainty in stochastic optimization problems [15], SP has been the most adopted, probably because of its flexibility in well-modeling a wide range of realistic applications [16–18]. The objective function of an SP problem typically involves computing an expected value. Unfortunately, in practice, it is difficult for a decision-maker to know exactly the probability distribution of all the random variables involved in the problem and to be able to analytically compute the multidimensional integral associated with the desired expected value. Therefore, in general, such an expected value is approximated through scenario generations, and the SP problem is substituted by a very large deterministic program in which second-stage variables and constraints are replicated for each considered scenario [19]. Since, in general, a consistent number of scenarios is needed for having a reasonable representation of the uncertainty, the computational burden of solving the relative deterministic program leads to the use of heuristic solution methods. Despite the practical utility of this approach, the introduction of such multiple levels of approximation makes the quality of the final solution somewhat questionable. For this reason, identifying and characterizing stochastic optimization problems for which it is possible to obtain optimal or approximated solutions without relying on scenario generation is a research question of paramount importance. This paper introduces and characterizes a large class of two-stage SP problems that can be solved to optimality by getting rid of scenario generation since a closed-form analytical expression of the expected value appearing in their objective function can be derived. Similar approaches have been developed for well-known optimization settings such as the *newsvendor problem* and other *Economic Order Quantity* models [20–22].

In this work, we focus on a particular class of two-stage Stochastic Programs that embed a set of DCPs. DCPs have been largely studied and applied in the decision theory literature, and many different models (e.g., logit- and probit-based models) have been proposed for solving them [23–25]. DCPs and the relative models have been embedded in a significant number of stochastic optimization problems to capture individuals' behaviors. A notable example is the so-called *multi-product*

Pricing Problem (mpPP) in which a seller is willing to compete within a market involving several customers and several alternative products. Assuming that a customer can only buy at most one of those products, the problem consists in finding the optimal prices at which each of the products should be sold such that the total revenue of the seller is maximized while satisfying market equilibrium constraints [26–28]. In these optimization problems, the demand for each product is typically proportional to the fraction of customers willing to choose it, and, in turn, such a fraction of customers depends on the probability that a customer will decide to buy a certain product rather than another. Since the total utility attributed to each product by each customer is stochastic, due to customers' idiosyncrasies, the setting can be generally approached as a DCP. In Du et al. [27], the authors considered a variant of the mpPP in which every single customer's utility for a given product is influenced by the overall interest of all customers for that product. Again, they formulate and solve the underlying optimization problem as a DCP, and a logit model is used to simulate customers' preferences. DCPs have also been embedded into SP formulations for modeling stochastic optimization problems affected by endogenous uncertainty. i.e., when its realizations are somehow dependent on the decision-maker's choices. Suitable examples can be found in Chen [29], as the problem of optimally allocating physicians to hospitals while accounting for patients' demand. Here, the probability that patients will require healthcare in a given hospital is determined by the distance between their residence and such hospital and the patient's preferences in terms of the perceived quality of the hospital services or a specific physician. Again, a DCP arises since the utility that a patient assigns to a given hospital is not fully known by the decision-maker. Note that, in this case, such a utility depends on the decisions regarding the assignment of physicians to hospitals. It is important to highlight that, despite their modeling power, SP problems embedding DCPs tend to be difficult to solve since even the easier behavioral models generally present non-linearities. In fact, the analytical expression of the choice probabilities is typically a non-linear and non-convex function of the stochastic utilities. Recently, Pacheco Paneque et al. [30] proposed a simulation-based framework to preserve the linearity and the convexity of the objective function of a mixed-integer program embedding DCPs. The proposed framework was successfully applied by Bortolomiol et al. [31,32] for finding equilibrium solutions within an oligopolistic transportation market and for addressing a pricing optimization problem. Note that, differently from the cited works, our tsSP-DCP framework provides a standard and easy-to-solve formulation for embedding, in a structured way, choice modeling within mathematical programming problems.

All the works listed in the previous paragraph relied on scenario generation-based approaches and various heuristic methods for solving stochastic problems embedding DCPs. However, a quite recent stream of research has shown that it is possible to leverage some classical Extreme Value Theory results [33,34] to derive an effective *Deterministic Approximation* (DA) of such problems. In particular, Tadei et al. [10,35] studied facility location and allocation problems, Perboli et al. [36] addressed a Knapsack Problem variant with multiple handlers in which a DCP is used for modeling the selection of the best handler for each loaded item, while Tadei et al. [37] and Fadda et al. [12] studied a multi-path Traveling Salesman Problem. A similar approach has also been used for approximating multi-stage decision processes [38,39] as well as a variant of the stochastic single machine scheduling problem [40]. However, the DA proposed in the above-mentioned literature shows two main drawbacks. First, the parameters characterizing the choice model of the various alternatives must be estimated to solve the obtained deterministic problem, and, therefore, the quality of the approximation strongly depends on the accuracy of such an estimation. Second, the DA derivation assumes that the stochastic utilities associated with the various alternatives of each DCP are independent and identically distributed, thus embracing only the simplest MEV choice model.[1] In this paper, instead, we formulate the necessary and sufficient assumptions to derive an exact analytical expression of the SP objective function involving the calculation of an expected value and, in turn, a scenario-independent deterministic equivalent form of all problems in the tsSP-DCP class. As a result, our framework is more general than what has appeared in the literature since the derived deterministic equivalent form also complies with cases in which the stochastic utilities associated with each DCP are not required to be independent of one another. Clearly, in the specific case where the stochastic utilities are independent, the closed-form analytical expression corresponds to the exact value to which the DA existing in the literature tends.

3. Two-stage stochastic programs embedding discrete choice problems

In Section 3.1, the mathematical structure of tsSP-DCPs is formally defined, while in Section 3.2, we focus on their interpretation in realistic decision-making settings and clarify their structure through peculiar examples.

3.1. Mathematical definition

The problems addressed in this paper, which we name tsSP-DCPs, are basically two-stage SP problems in which the second stage consists of a finite set of DCPs, each associated with one first-stage decision variable. In the following, we define their general form.

Let us consider:

- $\boldsymbol{\theta}$: a vector of random variables;
- $\boldsymbol{x}$, $\boldsymbol{y}$: the vector representing all the first-stage and second-stage decision variables, respectively;
- $f_1(\cdot)$, $f_2(\cdot)$: two scalar functions;
- $\mathcal{X}$, $\mathcal{Y}$: the feasibility set of the first-stage and of the second-stage decision variables, respectively.

Then, the general formulation of a two-stage SP problem can be written (see, e.g., [7]) as the following first-stage problem

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & f_1(\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{\theta}}[h(\boldsymbol{x},\boldsymbol{\theta})] \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{1}$$

where $h(\boldsymbol{x},\boldsymbol{\theta})$ is the optimum of the following second-stage problem

$$\begin{aligned} h(\boldsymbol{x},\boldsymbol{\theta}) \doteq \max_{\boldsymbol{y}} \quad & f_2(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta}) \\ \text{s.t.} \quad & \boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x},\boldsymbol{\theta}). \end{aligned} \tag{2}$$

The first stage tackles the uncertain profit regarding future decisions through its expected value. In the second stage, both the optimality and the feasibility of a solution may depend on the uncertainty and the decision made in the first stage. Note that, without loss of generality, we presented two-stage SP problems in their maximization form in line with discrete choice modeling, which primarily addresses problems aiming at the maximization of stochastic utilities.

We now formally define a DCP. Let us consider a decision-maker i and a set $\mathcal{L}_i$ containing a finite number of alternatives from which the decision-maker can choose. Let $U_{ij}(\theta_{ij})$ be the stochastic utility that the decision-maker attributes to alternative $j \in \mathcal{L}_i$. Without loss of generality, we define such a utility as the sum of two components, i.e.,

$$U_{ij}(\theta_{ij}) \doteq V_{ij} + \theta_{ij},$$

where V_{ij} is a deterministic value and θ_{ij} the corresponding stochastic oscillation. The deterministic part corresponds to the utility that

[1] Actually, in Fadda et al. [12], the authors were able to exploit the specific properties of the optimization setting at hand to partially relax such an assumption, thus asking only for asymptotical independence of the utility oscillations.

can be captured by any external observer of the problem, while the stochastic part corresponds to the idiosyncrasies of the decision-maker regarding such an alternative. Then, a DCP for the decision-maker i is the following stochastic optimization problem:

$$\begin{aligned} \max_{\boldsymbol{y}} \quad & \sum_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij}) y_{ij} \\ \text{s.t.} \quad & \sum_{j\in\mathcal{L}_i} y_{ij} = 1 \\ & y_{ij} \in \{0,1\}, \qquad \forall j \in \mathcal{L}_i. \end{aligned} \tag{3}$$

In formulation (3), y_{ij} is a binary variable taking value 1 if and only if the decision-maker i selects alternative $j \in \mathcal{L}_i$. The interpretation of a DCP is quite intuitive. In fact, the decision-maker wants to select exactly one alternative among many so as to maximize the utility deriving from the choice, given that an uncertain utility characterizes each alternative. However, note that in such settings, it is common to identify many decision-makers and a set $\mathcal{L}$ containing all the possible existing alternatives, being $\mathcal{L}_i \subseteq \mathcal{L}$ the subset of alternatives available for the decision-maker i.

Given the two previous stochastic optimization settings, we can now define the mathematical structure of a tsSP-DCP. Let us consider a set $\mathcal{I}$ of DCPs and a decision-maker for which:

- x_i is a first-stage decision variable different from 0 if and only if an alternative will be selected in the set $\mathcal{L}_i$ associated with the DCP $i \in \mathcal{I}$ at the second stage;
- y_{ij} is a second-stage binary decision variable equal to 1 if and only if the alternative $j \in \mathcal{L}_i$ is selected for the associated DCP $i \in \mathcal{I}$ at the second stage.

Then, a tsSP-DCP is the following two-stage SP problem:

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & f_1(\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{\theta}}[h(\boldsymbol{x},\boldsymbol{\theta})] \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{4}$$

where

$$\begin{aligned} h(\boldsymbol{x},\boldsymbol{\theta}) = \max_{\boldsymbol{y}} \quad & \sum_{i\in\mathcal{I}} x_i \sum_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij}) y_{ij} \\ \text{s.t.} \quad & \sum_{j\in\mathcal{L}_i} y_{ij} = \mathbb{1}_0(x_i), && \forall i \in \mathcal{I} \\ & y_{ij} \in \{0,1\}, && \forall j \in \mathcal{L}_i, i \in \mathcal{I}. \end{aligned} \tag{5}$$

with $\mathbb{1}_0(\cdot)$ being a function defined as

$$\mathbb{1}_0(z) = \begin{cases} 1 & \text{if } z \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

We remark that the second stage of a tsSP-DCP has a precise and quite simple combinatorial structure. In particular, exactly one alternative $j \in \mathcal{L}_i$ must be selected for each DCP $i \in \mathcal{I}$ associated with a non-null first-stage variable x_i, and there are no interdependencies between the decisions made within each DCP and those made within the others. Instead, the first stage complies with every possible requirement, represented by the feasibility set $\mathcal{X}$. Moreover, in a tsSP-DCP, uncertainty only affects the second-stage objective function and not its feasibility set. Finally, second-stage utilities do not explicitly depend on the first-stage decisions, but only on the uncertainty.

A notable tsSP-DCP special case, especially important for modeling several location and service network design problems, arises when all the first-stage decisions are binary (i.e., when $x_i \in \{0,1\}, \forall i \in \mathcal{I}$). In this case, $\mathbb{1}_0(x_i) = x_i$ for each $i \in \mathcal{I}$ and the second stage in (5) becomes

$$\begin{aligned} h(\boldsymbol{x},\boldsymbol{\theta}) = \max_{\boldsymbol{y}} \quad & \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij}) y_{ij} \\ \text{s.t.} \quad & \sum_{j\in\mathcal{L}_i} y_{ij} = x_i, && \forall i \in \mathcal{I} \\ & y_{ij} \in \{0,1\}, && \forall j \in \mathcal{L}_i, i \in \mathcal{I}, \end{aligned} \tag{7}$$

in which the second-stage objective function has become linear, although depending on a stochastic parameter.

3.2. Exemplification and interpretation

In the following, we illustrate the introduced tsSP-DCPs and their characteristics through some concrete examples, demonstrating how this class spans a very large number of applications in different relevant Management Science and Operations Research fields (logistics and transportation, facility location, scheduling).

Let us consider a company that faces the here-and-now problem of selecting a portfolio $\mathcal{I}$ of potential facilities to locate (e.g., projects to invest in, physical services to install), and later the problem of choosing the appropriate assignment of each selected facility $i \in \mathcal{I}$ to one alternative (e.g., operators of a project, handlers or customers of a service) among a set of alternatives $\mathcal{L}_i$. The location phase may generally be complicated by specific requirements such as budget constraints or incompatibility between facilities. For simplicity, let us only consider a budget constraint represented by a maximum number p of facilities to locate. Moreover, each located facility $i \in \mathcal{I}$ yields a fixed opening cost c_i and a different profit depending on the later assignment (e.g., the profit of a project will depend on the operator that is going to manage it). Profits are not deterministically known at the time of the facility location phase (e.g., they may rely on a bidding auction among the available operators that takes place later in time). Hence, we represent the uncertain profit as a utility $U_{ij}(\theta_{ij}) = V_{ij} + \theta_{ij}$, where V_{ij} and θ_{ij} are the deterministic profit and the stochastic profit oscillation associated with the assignment of alternative $j \in \mathcal{L}_i$ to the located facility $i \in \mathcal{I}$. The above setting can be modeled as a stochastic facility location problem aiming at deciding the p facilities to locate and the appropriate assignment of an alternative to each selected facility so as to maximize the expected total profit, net of the location cost. A stochastic integer linear programming formulation of this problem is

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & -\sum_{i\in\mathcal{I}} c_i x_i + \mathbb{E}_{\boldsymbol{\theta}}\left[\max_{\boldsymbol{y}} \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{L}_i} (V_{ij} + \theta_{ij}) y_{ij}\right] \\ \text{s.t.} \quad & \sum_{i\in\mathcal{I}} x_i \leq p \\ & \sum_{j\in\mathcal{L}_i} y_{ij} \leq x_i, && \forall i \in \mathcal{I} \\ & x_i \in \{0,1\}, && \forall i \in \mathcal{I} \\ & y_{ij} \in \{0,1\}, && \forall i \in \mathcal{I}, j \in \mathcal{L}_i, \end{aligned} \tag{8}$$

where x_i is a binary first-stage variable equal to 1 if facility $i \in \mathcal{I}$ is located, and 0 otherwise, and y_{ij} is a binary second-stage variable (thus depending on the uncertainty realization) equal to 1 if alternative $j \in \mathcal{L}_i$ is assigned to the located facility $i \in \mathcal{I}$, and 0 otherwise. The objective function maximizes the expected total assignment profit minus the total location cost. In the absence of full information on the stochastic profits at the first stage, it is reasonable for the company to select the facilities to invest in by maximizing the total expected profit that would return from the assignment at the second stage, minus the fixed opening costs. The two sets of structural constraints ensure that a maximum of p facilities are located and that an alternative is assigned to a facility only if the latter has been located, respectively. The obtained model is a tsSP-DCP. In fact, each facility chosen at the first stage determines the need to solve a DCP at the second stage, i.e., to choose the most profitable assignment at the second stage among different alternatives.

Several other examples of stochastic optimization problems whose SP formulation belongs to the tsSP-DCP class can be directly found in the recent literature. In Tadei et al. [37] and Fadda et al. [12], the authors addressed the so-called *multi-path Traveling Salesman Problem* (mpTSP), a generalization of the classical TSP in which alternative paths are available between two different nodes. Those paths might represent different transport modes or different physical routes, and each alternative is characterized by a stochastic cost whose probability distribution is unknown. The problem consists in finding a Hamiltonian tour of the nodes and, after the realization of the uncertainty, the path to be used between any pair of locations to minimize the total expected cost of the tour. The mpTSP can be modeled as a tsSP-DCP since, given any pair of consecutive nodes to visit in the optimal Hamiltonian tour, the decision-maker needs to solve a DCP concerning the selection of one path among the available ones, and each DCP is triggered by the first-stage decision of including the relative arc in the optimal solution. Instead, in the *multi-handler Knapsack Problem* introduced by Perboli et al. [36], a set of potential handlers needs to be chosen for managing a set of items, each guaranteeing a stochastic utility depending on the handler that manages it. The problem consists in finding the optimal subset of items to be loaded in the knapsack and, after the realization of the uncertainty, determining the handler that is going to process each loaded item so as to maximize the total utility while satisfying the knapsack capacity constraint. This generalization of the Knapsack Problem, arising in many logistics applications, embeds a DCP aiming at selecting the best handler for each item loaded in the knapsack at the first stage. Finally, Roohnavazfar et al. [40] introduced a stochastic variant of the single-machine multi-configuration scheduling problem, which belongs again to the tsSP-DCP class. In such a problem, a DCP arises in the selection of the optimal configuration under which the machine should process the next scheduled job, and the scheduling decisions at the first stage trigger several DCPs in the second stage, each one independent of the others.

In general, at the first stage of a tsSP-DCP, the decision-maker acts on some first-stage decision variables that can *trigger* second-stage DCPs. In the second stage, instead, he has to solve all the triggered DCPs. In the case of the previous example, each chosen project at the first stage triggers a DCP aiming at optimally choosing an operator that is going to manage it. It is important to notice that, even if it involves many different DCPs, a tsSP-DCP does not assume the existence of different decision-makers acting in the same micro-economic setting, as in Game Theory. Instead, in a tsSP-DCP, we can interpret any DCP (i.e., any $i \in \mathcal{I}$) as a specific micro-economic scenario in which a single decision-maker is going to operate. The chance or the weight of each scenario within the overall problem depends on the decision made at the first stage regarding the corresponding variable x_i. In fact, while all the above examples involve pure binary decisions at the first stage, the tsSP-DCP definition allows for more general integer or continuous variables for which the decision-maker must attribute a certain level of activity.

4. Deterministic equivalent form of MEV-based tsSP-DCPs

In Section 4.2, we derive a deterministic program that results in being equivalent to the tsSP-DCP and is based on a closed-form analytical expression of the second-stage expected optimum. Then, in Section 4.3, we clarify how the derived results can be widely applied and turn out to be useful in practice. Since our derivation holds for DCPs following MEV choice models, we formally discuss the basic assumptions on which MEV models rely (Section 4.1).

4.1. MEV models

To fully understand the theoretical results obtained, we recall and briefly discuss the two conditions underlying the MEV modeling framework.

The first condition assumes that each random vector of oscillations $[\theta_{i,1},\dots,\theta_{i,|\mathcal{L}_i|}]^T$, for each DCP $i \in \mathcal{I}$, follows a multivariate extreme value distribution. More formally:

Assumption 1. For any $i \in \mathcal{I}$, the joint cumulative probability distribution $F_i(z_1,\dots,z_{|\mathcal{L}_i|})$ of each random oscillations vector $[\theta_{i,1},\dots,\theta_{i,|\mathcal{L}_i|}]^T$ is such that there exists a differentiable function $G_i : \mathbb{R}_+^{|\mathcal{L}_i|} \to \mathbb{R}_+$ (called the **generating function**) for which it holds

$$F_i(z_1,\dots,z_{|\mathcal{L}_i|}) = \exp\left(-G_i\left(e^{-z_1},\dots,e^{-z_{|\mathcal{L}_i|}}\right)\right). \tag{9}$$

For F_i to be a proper joint cumulative probability distribution, it must hold that

$$\lim_{z_j\to-\infty} F_i(z_1,\dots,z_{|\mathcal{L}_i|}) = 0,\ \forall j \in \mathcal{L}_i, \tag{10}$$

$$\lim_{z_1\to+\infty,\dots,z_{|\mathcal{L}_i|}\to+\infty} F_i(z_1,\dots,z_{|\mathcal{L}_i|}) = 1, \tag{11}$$

and that any of its partial derivatives of any order must be positive, i.e., by denoting as I_k any set of k indexes chosen in $\mathcal{L}_i$, it must hold

$$\frac{\partial^k F_i}{\partial z_{I_1},\dots,\partial z_{I_k}}(z_1,\dots,z_{|\mathcal{L}_i|}) \geq 0, \quad \forall (z_1,\dots,z_{|\mathcal{L}_i|}) \in \mathbb{R}^{|\mathcal{L}_i|},\ \forall k = 1,\dots,|\mathcal{L}_i|. \tag{12}$$

According to Assumption 1, properties (10), (11), and (12) respectively imply for G_i that

$$\lim_{z_j\to+\infty} G_i(z_1,\dots,z_{|\mathcal{L}_i|}) = +\infty,\ \forall j \in \mathcal{L}_i, \tag{13}$$

$$\lim_{z_1\to0,\dots,z_{|\mathcal{L}_i|}\to0} G_i(z_1,\dots,z_{|\mathcal{L}_i|}) = 0, \tag{14}$$

and that its partial derivatives alternate their sign, i.e.,

$$(-1)^{k-1}\frac{\partial^k G_i}{\partial z_{I_1},\dots,\partial z_{I_k}}(z_1,\dots,z_{|\mathcal{L}_i|}) \geq 0, \quad \forall (z_1,\dots,z_{|\mathcal{L}_i|}) \in \mathbb{R}_+^{|\mathcal{L}_i|},\ \forall k = 1,\dots,|\mathcal{L}_i|. \tag{15}$$

It is important to remark that the generating function's properties reported in (13)–(15) only ensure that the function F_i in (9) is a proper joint cumulative probability distribution. As such, they are implicit in Assumption 1 and must not be considered additional requirements.

The second condition assumes that the generating function satisfies the so-called *homogeneity property* with respect to a constant β. This is formally expressed as follows.

Assumption 2. For any $i \in \mathcal{I}$, the function G_i is homogeneous of degree β, or β–**homogeneous**, i.e., it exists a constant $\beta > 0$ such that it holds

$$G_i\left(\alpha(z_1,\dots,z_{|\mathcal{L}_i|})\right) = \alpha^\beta G_i(z_1,\dots,z_{|\mathcal{L}_i|}) \tag{16}$$

for all $(z_1,\dots,z_{|\mathcal{L}_i|}) \in \mathbb{R}_+^{|\mathcal{L}_i|}$ and any real value $\alpha > 0$.

Regarding β–homogeneous G_i functions, we state the following three important properties, which will be useful for proving our main result.

Property 1. *Function G_i is β–homogeneous if and only if, for each $j = 1,\dots,|\mathcal{L}_i|$, it exists a constant $\beta > 0$ such that*

$$G_i(z_1,\dots,z_{|\mathcal{L}_i|}) = z_j^\beta G_i\left(\frac{z_1}{z_j},\dots,1,\dots\frac{z_{|\mathcal{L}_i|}}{z_j}\right),\ \forall (z_1,\dots,z_{|\mathcal{L}_i|}) \in \mathbb{R}_+^{|\mathcal{L}_i|}.$$

Property 2 (*Euler's theorem*)**.** *Function G_i is β–homogeneous if and only if*

$$\sum_{j=1}^{|\mathcal{L}_i|} z_j \frac{\partial G_i(z_1,\dots,z_{|\mathcal{L}_i|})}{\partial z_j} = \beta G_i(z_1,\dots,z_{|\mathcal{L}_i|}),\ \forall (z_1,\dots,z_{|\mathcal{L}_i|}) \in \mathbb{R}_+^{|\mathcal{L}_i|}.$$

Property 3. *If function G_i is β–homogeneous, then its derivative $\frac{\partial G_i}{\partial z_j}$ is $(\beta-1)$–homogeneous for each $j = 1,\dots,|\mathcal{L}_i|$.*

The two assumptions introduced lead to the derivation of a broad class of well-studied probabilistic choice frameworks, namely the MEV models [41–43], which translate the theory of rational choice behavior into concrete mathematical tools that enable empirical analyses. In Assumption 1, the generating function G_i is introduced to model possible dependencies among the utilities of the available alternatives for DCP $i \in \mathcal{I}$, thus generalizing the summation operator used in the classical logit formulation.[2] This provides a flexible mechanism to account for the correlation generated by alternatives that exhibit similarities or overlapping attributes. When Assumption 2 also holds, it is possible to show that any marginal distribution of F_i is a (univariate) extreme value distribution in the form

$$F_i(z_j) = \exp\left(e^{-\beta z_j}\right), \ \forall j \in \mathcal{L}_i$$

and that the corresponding choice model has a closed-form analytical expression, i.e.,

$$\mathbb{P}[\text{selecting } j \in \mathcal{L}_i] = \frac{e^{V_{ij}} \frac{\partial G}{\partial z_i}\left(e^{V_{i,1}}, \dots, e^{V_{i,|\mathcal{L}_i|}}\right)}{\beta G\left(e^{V_{i,1}}, \dots, e^{V_{i,|\mathcal{L}_i|}}\right)}. \tag{17}$$

The choice model in (17) is consistent with the random utility maximization theory [13,44], implying rational preferences for the decision-maker (complete and transitive) and choice probabilities that satisfy economic properties such as monotonicity and regularity. For example, by using Property 3, we can see how (17) implies that if the deterministic utility of all the alternatives is augmented by a constant factor, then the probability of selecting a specific alternative does not change.

4.2. Deterministic equivalent form

We are now ready to present the main result of this work, i.e., the derivation of a deterministic program totally equivalent to a tsSP-DCP. This is stated in the following theorem.

Theorem 1. *Consider any tsSP-DCP as defined in (4)–(5). If the decision-maker's choice behavior can be modeled by an MEV, then the optimal solution of the stochastic program in (4) can be found by solving the following deterministic program*

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & f_1(\boldsymbol{x}) + \sum_{i\in\mathcal{I}} \left[\frac{1}{\beta}\left(\ln\left(G_i\left(e^{V_{i,1}}, \dots, e^{V_{i,|\mathcal{L}_i|}}\right)\right) + \gamma\right)\right] x_i \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{18}$$

where G_i is the generating function of the DCP $i \in \mathcal{I}$ satisfying Assumptions 1 and 2, β is the homogeneity degree of the function G_i, V_{ij} is the deterministic utility associated with alternative $j \in \mathcal{L}_i$ within the DCP $i \in \mathcal{I}$, and $\gamma = -\int_0^{+\infty} \ln(t)e^{-t}dt$ is the Euler constant.

Proof. In any tsSP-DCP (4)–(5), the second-stage problem (5) has the same optimal solution value of

$$\begin{aligned} \max_{\boldsymbol{y}} \quad & \sum_{i\in\mathcal{I}} x_i \mathbb{1}_0(x_i) \sum_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij}) y_{ij} \\ \text{s.t.} \quad & \sum_{j\in\mathcal{L}_i} y_{ij} = 1, && \forall i \in \mathcal{I} \\ & y_{ij} \in \{0,1\}, && \forall j \in \mathcal{L}_i, i \in \mathcal{I}, \end{aligned} \tag{19}$$

which can be equivalently rewritten as

$$\begin{aligned} \max_{\boldsymbol{y}} \quad & \sum_{i\in\mathcal{I}} x_i \sum_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij}) y_{ij} \\ \text{s.t.} \quad & \sum_{j\in\mathcal{L}_i} y_{ij} = 1, && \forall i \in \mathcal{I} \\ & y_{ij} \in \{0,1\}, && \forall j \in \mathcal{L}_i, i \in \mathcal{I}, \end{aligned} \tag{20}$$

given the definition of $\mathbb{1}_0(\cdot)$ as proposed in (6).

From formulation (20), it is evident that the second-stage problem can be decomposed into $|\mathcal{I}|$ independent DCPs, as those defined in (3), each having an optimal objective value multiplied by the relative first-stage variable x_i. Hence, the second-stage optimal solution can be obtained by selecting, for each set of alternatives $\mathcal{L}_i$, the one with the largest utility. Let us denote as j_i^* the alternative associated with the largest utility in the set $\mathcal{L}_i$, after the resolution of the uncertainty on the utilities. Then, the second-stage optimal solution $\boldsymbol{y}^*$ can be found by setting

$$y_{ij}^* = \begin{cases} 1 & \text{if } j = j_i^* \\ 0 & \text{otherwise,} \end{cases} \quad i \in \mathcal{I}. \tag{21}$$

This means that the expected value of the second-stage problem is

$$\begin{aligned} \mathbb{E}_\theta[h(\boldsymbol{x},\boldsymbol{\theta})] &= \mathbb{E}_\theta\left[\max_{y_{ij}\in\{0,1\}, j\in\mathcal{L}_i, i\in\mathcal{I}} \sum_{i\in\mathcal{I}} x_i \sum_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij}) y_{ij} \ \middle|\ \sum_{j\in\mathcal{L}_i} y_{ij} = 1, \forall i \in \mathcal{I}\right] = \\ &= \mathbb{E}_\theta\left[\sum_{i\in\mathcal{I}} U_{i,j_i^*}(\theta_{i,j_i^*}) x_i\right] = \\ &= \mathbb{E}_\theta\left[\sum_{i\in\mathcal{I}} \left(\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right) x_i\right]. \end{aligned}$$

Then, by leveraging the linearity of the expected value, we obtain

$$\mathbb{E}_\theta[h(\boldsymbol{x},\boldsymbol{\theta})] = \mathbb{E}_\theta\left[\sum_{i\in\mathcal{I}} \left(\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right) x_i\right] = \sum_{i\in\mathcal{I}} \mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] x_i. \tag{22}$$

Now, our tsSP-DCP (4)–(5) has been reduced to the following formulation

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & f_1(\boldsymbol{x}) + \sum_{i\in\mathcal{I}} \mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] x_i \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{23}$$

which is still a stochastic program due to the stochastic profit oscillations θ_{ij}. The objective function in (23) highlights that the expected value of the second-stage problem of a tsSP-DCP corresponds to a function, linear in $\boldsymbol{x}$, of $|I|$ independent expected maximum utilities.

In Discrete Choice theory, i.e., where a rational decision-maker, after assigning a stochastic utility to each alternative, chooses the one with the highest utility, typically happens to look for the expectation of the maximum of a sequence of random variables, as in this case. Since such a choice behavior is assumed to be captured by an MEV model, it is possible to derive a closed-form analytical expression for the expected maximum utility of each DCP $i \in \mathcal{I}$ resulting from the choice performed in the set $\mathcal{L}_i$. More precisely, using the calculus-based derivation presented in Theorem 2 (see Appendix), we obtain

$$\mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] = \frac{1}{\beta}\left(\ln\left(G_i\left(e^{V_{i,1}}, \dots, e^{V_{i,|\mathcal{L}_i|}}\right)\right) + \gamma\right), \ i \in \mathcal{I}. \tag{24}$$

Now, the theorem follows by plugging (24) into (23). □

We name the program in (18) as the *MEV-based deterministic equivalent form* of a tsSP-DCP. Such a formulation, apart from being deterministic, has very interesting advantages from a computational point of view. In fact, since the analytical expression of the second-stage expected value in the objective function only requires a function evaluation, problem (18) has the same computational complexity as the relative tsSP-DCP under no uncertainty (i.e., the problem in which stochastic parameters are substituted by deterministic values). For example, if the feasibility set $\mathcal{X}$ of a tsSP-DCP can be represented by linear constraints and the function f_1 is linear in $\boldsymbol{x}$, then the obtained deterministic equivalent form in (18) results to be a linear program as well. Moreover, differently from classical scenario-based deterministic equivalent formulations, the dimensions in terms of the number of variables and constraints of problem (18) are the same as those of the original tsSP-DCP first-stage.

Note that the derived deterministic program explicitly allows computing only the first-stage decision variables of the original tsSP-DCP, given that the second-stage decisions are implicitly addressed by the approach. However, this does not represent a limitation for two reasons. First, in two-stage SP problems, the uncertainty affects only the first-stage decision variables since the computation of the second-stage ones

[2] Note that univariate distributions although enriched with a general generating function are commonly named Generalized Extreme Value (GEV) distributions.

occurs only after the realization of all random parameters. Second, the second-stage variables will be decided when the uncertainty is revealed according to the utility maximization criterion and independently for each second-stage DCP (so, without the actual need for a mathematical program).

4.3. Applicability of the framework

We first describe some notable MEV models, then we create a direct link between this work and previous approaches in which particular cases have been addressed for solving realistic problems.

4.3.1. Notable MEV models

In the following, we show that the two formal assumptions (introduced in Section 4.1) needed to obtain MEV choice models are not as restrictive as may seem. Instead, MEVs are a large family of statistical models, particularly interesting in practice and commonly used in choice modeling. Three well-known MEVs are briefly described in the following, where a single DCP facing a set $\mathcal{L}$ of alternatives is given. Each model derives from a specific choice of the generating function G.

First, the generating function

$$G(z_1, \dots, z_{|\mathcal{L}|}) \doteq \sum_{j\in\mathcal{L}} z_j^{\beta} \tag{25}$$

verifies the assumptions. It amounts to assuming that the random oscillations are independent of each other, thus the vector θ is a collection of independent univariate extreme value distributions. In the context of choice modeling, this corresponds to the well-known *logit* model.

Assume now that the set of alternatives $\mathcal{L}$ is partitioned into r sets C_m called *nests*, i.e., $\bigcup_{m=1}^{r} C_m = \mathcal{L}$ and $C_{m'} \cap C_{m''} = \emptyset, \forall m', m'' = 1, \dots, r, m' \neq m''$. The generating function

$$G(z_1, \dots, z_{|\mathcal{L}|}) \doteq \sum_{m=1}^{r} \left(\sum_{j\in C_m} z_j^{\beta_m} \right)^{\beta/\beta_m} \tag{26}$$

verifies the assumptions if $0 < \beta \leq \beta_m$. If j_1 and j_2 both belong to nest C_m, it can be shown that the corresponding oscillations have a correlation equal to $\rho_{j_1,j_2} = 1 - \left(\frac{\beta}{\beta_m}\right)^2$, otherwise they are independent. In choice modeling, this corresponds to the so-called *nested logit* model.

Finally, consider a series of r nests C_m for which it only holds that $\bigcup_{m=1}^{r} C_m = \mathcal{L}$. In this case, since a partition of the set $\mathcal{L}$ is not imposed, an alternative may belong to several nests. The generating function

$$G(z_1, \dots, z_{|\mathcal{L}|}) \doteq \sum_{m=1}^{r} \left(\sum_{j\in\mathcal{L}} \alpha_{jm}^{\frac{\beta_m}{\beta}} z_j^{\beta_m} \right)^{\beta/\beta_m} \tag{27}$$

verifies the assumptions if $0 < \beta \leq \beta_m$, $\sum_{j\in\mathcal{L}} \alpha_{jm} > 0$ for all $m = 1, \dots, r$, and $\alpha_{jm} \geq 0$ for all $m = 1, \dots, r$ and for all $j \in \mathcal{L}$ [45]. In the context of choice modeling, this corresponds to the so-called *cross-nested logit* model. Such a function is very flexible and generalizes both the logit and the nested-logit models. Indeed, it can be shown that any random utility choice model can be approximated arbitrarily well through a cross-nested logit model [14, Theorem 4].

4.3.2. Heuristic usage of the deterministic equivalent form

As mentioned in the literature review of Section 2, some previous papers have already focused on different applications (spanning from knapsack-like to vehicle routing and production scheduling problems) that can be modeled as tsSP-DCPs [10,12,36,39]. The approach followed in these works, which leads to a deterministic equivalent form that only approximates the optimal solution of a tsSP-DCP, can be seen as a heuristic usage of the exact results we are proposing here. More precisely, in all these works, each random oscillation is assumed to be the maximum or the minimum of a sequence of other random variables whose actual distribution is unknown. This means that, as commonly happens in practice, Assumptions 1 and 2 do not hold or cannot be verified. Moreover, the analytical expression derived in (A.1) cannot even be computed since both $G_i, \forall i \in \mathcal{I}$, and β are not known. However, under mild assumptions on the unknown distribution of the random variables that constitute the sequences (e.g., a double exponential decrease of the distribution tails) and through an ad-hoc calibration of the distribution parameters (β in particular), it is possible to show that each random oscillation can be considered to be following a univariate extreme value distribution as the sequence length tends to infinity [46]. Since, as already noted in Section 4.1, this is equivalent to assuming a β–homogeneous multivariate extreme value distribution for the random oscillations [13], the studied tsSP-DCPs can be deterministically approximated when the number of scenarios is sufficiently large. Note that, in all these works, the authors have assumed independent random oscillations, which is equivalent to assuming a generating function as that stated in (25) and leading to a logit choice model. Only in Fadda et al. [12], given some equilibrium properties of the realistic traffic network underlying the considered routing problem, it was possible to slightly relax such an assumption and ask the random variables to be just asymptotically independent.

The above-described works demonstrate that, even when the underlying assumptions do not hold and thus our exact analytical result cannot be achieved, it is possible to use our framework effectively in a heuristic way for obtaining a good estimation of the considered tsSP-DCP when the embedded DCPs are described by a suitable choice model and the related generating function's parameters are well-calibrated. Interestingly, the experiments presented have empirically shown that the estimator of the second-stage expected optimum of a tsSP-DCP obtained under the assumption that the random oscillations follow an extreme value distribution still performs well in practice even when the uncertain values actually behave differently. This means that our approach has an enormous potential for applicability in many realistic settings, where it is common to face stochastic data for which the probability distributions are totally unknown or just partially estimable.

5. Experimental analysis

In this section, we validate the efficiency, quality, and robustness of our approach through computational experiments on a showcase tsSP-DCP. The showcase is presented in Section 5.1, the relative instance generation in Section 5.2, and the discussion of the results obtained in Section 5.3.

The proposed method is implemented in Python 3, and all the models are solved by Gurobi v9.1.1 via its Python APIs. The Euler constant γ, needed to derive the closed-form analytical expression, is approximated to its tenth decimal digit. All the tests have been done on an *Intel(R) Core(TM)* i7-5500U CPU@2.40 GHz computer with 16 GB of RAM and running *Ubuntu* v22.04.

5.1. Showcase

As a showcase tsSP-DCP, we consider the facility location problem with stochastic allocation profits already discussed in Section 3.2 and formulated in (8). For this problem, we assume to face stochastic allocation profit vectors $[\theta_{i,1}, \dots, \theta_{i,|\mathcal{L}_i|}]^T$, for each facility $i \in \mathcal{I}$, which are generated by joint probability distributions complying with Assumptions 1 and 2. The application of our method to this simple setting, instead of focusing on more computationally challenging problems, enables us to get useful insights, thus getting rid of too many external factors.

A typical scenario-based Deterministic Equivalent Problem (DEP^{SB}) of (8) can be obtained by discretizing the underlying probability distribution, i.e., by defining a finite set $\mathcal{S}$ of scenarios and by obtaining, for each scenario $s \in \mathcal{S}$, a probability π^s to occur and a deterministic realization θ_{ij}^s of the uncertain parameter. The DEP^{SB} of problem (8) is then

$$(DEP^{SB})\quad \max_{\mathbf{x},\mathbf{y}} \quad -\sum_{i\in\mathcal{I}} c_i x_i + \sum_{s\in\mathcal{S}} \pi^s \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{L}_i} (V_{ij} + \theta_{ij}^s) y_{ij}^s$$
$$\text{s.t.} \quad \sum_{i\in\mathcal{I}} x_i \le p$$
$$\sum_{j\in\mathcal{L}_i} y_{ij}^s \le x_i, \qquad \forall i \in \mathcal{I}, s \in \mathcal{S}$$
$$x_i \in \{0,1\}, \qquad \forall i \in \mathcal{I}$$
$$y_{ij}^s \in \{0,1\}, \qquad \forall i \in \mathcal{I}, j \in \mathcal{L}_i, s \in \mathcal{S} \tag{28}$$

where y_{ij}^s is a binary variable equal to 1 if alternative $j \in \mathcal{L}_i$ is assigned to the located facility $i \in \mathcal{I}$ in scenario $s \in \mathcal{S}$, and 0 otherwise.

However, since the generating function G_i is assumed to be known for each $i \in \mathcal{I}$ as well as the relative homogeneity degree β, we can also obtain the MEV-based Deterministic Equivalent Problem (DEP^{MEV}) of (8) by directly applying Theorem 1, i.e.,

$$(DEP^{MEV})\quad \max_{\mathbf{x}} \quad \sum_{i\in\mathcal{I}} \left[\frac{1}{\beta}\left(\ln\left(G_i\left(e^{V_{i,1}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\right)+\gamma\right) - c_i\right] x_i$$
$$\text{s.t.} \quad \sum_{i\in\mathcal{I}} x_i \le p$$
$$x_i \in \{0,1\}, \quad \forall \in \mathcal{I}. \tag{29}$$

From now on, we will call $\mathbf{x}^{SB}$ and $\mathbf{x}^{MEV}$ the (first-stage) optimal solution of DEP^{SB} and of DEP^{MEV}, respectively.

5.2. Instance generation and simulation of the uncertainty

In the following, we describe the generation process of the instances used to test the proposed showcase tsSP-DCP. Our goal is not to generate ad-hoc data complying with some realistic industrial setting, but to create instances as general and scalable as possible to represent many different situations. All the scripts have been implemented using the *R* language and are available, together with the generated instances, at https://github.com/EdoF90/EVT-DEF.

For all the instances, we considered $|\mathcal{I}| = 100$ potential facilities and $|\mathcal{L}_i| = 10$ alternatives $\forall i \in \mathcal{I}$. We generated the deterministic fixed cost c_i to locate facility $i \in \mathcal{I}$ by uniformly sampling a real number in the interval (0, 0.2), while the maximum number of facilities to locate p has been set to a certain percentage of $|\mathcal{I}|$, namely 5%, 10%, 20%, and 25%. Note that p is always a small proportion of $|\mathcal{I}|$ to allow the existence of many different possible first-stage feasible solution vectors and, in turn, the possibility to obtain improvements by selecting different facilities through different methods.

The stochastic profits are sampled from two probability distributions, thus obtaining two classes of instances:

- **logit instances (L)**: given a facility $i \in \mathcal{I}$, all the related random variables $\theta_{ij}, j \in \mathcal{L}_i$, are assumed to be independent. Hence, the generating function G_i of the joint cumulative probability distribution of the random variables is the one defined in Eq. (25), which corresponds to the well-known logit model in choice theory. Then, the realizations θ_{ij}^s for each scenario $s \in \mathcal{S}$ of each random variable θ_{ij} can be simply obtained by sampling from an extreme value (Gumbel) distribution. We set the location parameter to 0 and randomly select the scale parameter (which is β) in the range $[1,2]$.
- **nested logit instances (NL)**: the random variables $\theta_{ij}, j \in \mathcal{L}_i$, associated with the facility $i \in \mathcal{I}$ are assumed to be dependent, and their interactions are simulated by choosing Eq. (26) as the generating function of the joint cumulative distribution of these random variables. This corresponds to the well-known nested logit model in choice theory. Obtaining the realizations θ_{ij}^s for each scenario $s \in \mathcal{S}$ of each random variable θ_{ij} is not as trivial as in the previous case. In fact, we would like to sample realizations from a joint cumulative distribution which, for each facility $i \in \mathcal{I}$, is

$$F_i(z_1,\dots,z_{|\mathcal{L}_i|}) = \exp\left(-\sum_{m=1}^{r}\left(\sum_{j\in\mathcal{C}_m} e^{-\beta_m z_j}\right)^{1/\beta_m}\right), \tag{30}$$

where $\mathcal{C}_m$ represents the mth nest of alternatives, and $\bigcup_{m=1}^{r} \mathcal{C}_m = \mathcal{L}_i$. In order to obtain (30), as commonly done in the nested logit model literature, the homogeneity parameter β appearing in the general function (26) has been normalized to 1. This way, only the β_m parameters must be set. Unfortunately, as far as we know, there are no ready-to-use tools able to sample from such a distribution. However, the *R* package[3] developed by [47] can simulate realizations from a multivariate extreme value distribution in the form

$$\tilde{F}_i(z_1,\dots,z_{|\mathcal{L}_i|}) = \exp\left(-\sum_{m=1}^{r}\left(\sum_{j\in\mathcal{C}_m} (1-z_j)^{\beta_m}\right)^{1/\beta_m}\right). \tag{31}$$

By observing that $F_i(z_1,\dots,z_{|\mathcal{L}_i|}) = \tilde{F}_i\left(1-e^{-z_1},\dots,1-e^{-z_{|\mathcal{L}_i|}}\right)$, it is possible to obtain realizations θ_{ij}^s of the random variables θ_{ij} by first sampling the realizations $\tilde{\theta}_{ij}^s, j \in \mathcal{L}_i$ from $\tilde{F}_i(z_1,\dots,z_{|\mathcal{L}_i|})$ and then calculate the needed realizations as $\theta_{ij}^s = -\ln(1-\tilde{\theta}_{ij}^s)$. When simulating from (31), the value of the parameter β_m for each nest $m = 1,\dots,M$ has been uniformly sampled in the interval $[2,5]$.

Finally, for both L and NL instances, we generated the deterministic utilities $V_{ij}, i \in \mathcal{I}, j \in \mathcal{L}_i$, by uniformly sampling their values in the range $[0,\bar{V}]$, where $\bar{V}$ depends on the average expected value (μ) of all the probability distributions of the random variables involved and on a parameter ν that allows us to control the amount of data stochasticity. More precisely,

$$\bar{V} = 2\mu\frac{100-\nu}{\nu}$$

with

$$\mu = \frac{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{L}_i}\mathbb{E}[\theta_{ij}]}{\sum_{i\in\mathcal{I}}|\mathcal{L}_i|} \quad \text{and} \quad 0 < \nu \le 100.$$

Basically, the larger the value of ν, the smaller the proportion of the deterministic utilities with respect to their stochastic counterpart.

For the experiments discussed in the next section, we generated 10 random repetitions for the two classes of instances (L or NL) and 9 different values of ν (from 10 to 90, with step 10). For each of the above generated stochastic settings, the problem is evaluated against 4 values of p, namely, $p = \{5,10,20,25\}$. This means a total of 720 benchmark instances. It is important to notice that modern solvers can tackle facility location problems of such dimensions quite efficiently, even when considering many scenarios. However, the use of this showcase allows us to experimentally validate on a large number of instances solved to optimality the proportional advantage of using our method in terms of solution quality and CPU times.

5.3. Results and discussion

In the following, we present and discuss the results obtained from different experiments of interest. Briefly, such experiments seek to answer the following questions (each addressed in a specific subsection):

- Is it worth addressing the tsSP-DCP at hand through a complex SP environment instead of just approximating the random variables by their expected value estimators?

[3] https://cran.r-project.org/web/packages/evd/index.html. Last accessed: 27-02-2025.

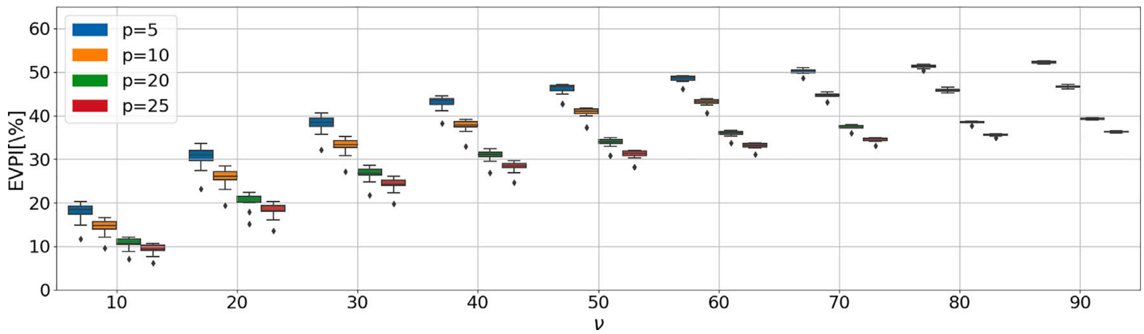

Fig. 1. $EVPI[\%]$ against different values of p and ν for L instances.

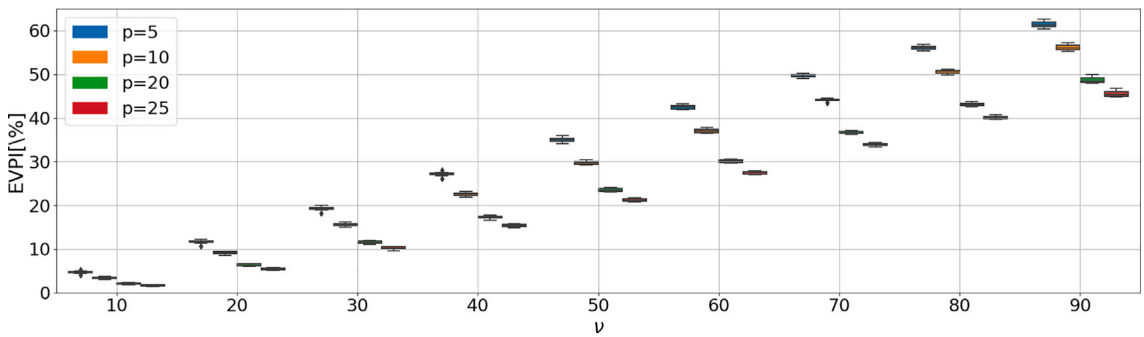

Fig. 2. $EVPI[\%]$ against different values of p and ν for NL instances.

- How many scenarios would be necessary for a scenario-based DEP to achieve a solution exhibiting, with respect to the optimal solution returned by our framework, a relative error that does not exceed a given threshold? Given such a number of scenarios, how much more computationally demanding is it to solve the scenario-based DEP with respect to ours?
- What would be the expected relative error of a solution obtained from our method if one does not know the exact distribution of the random data but has just a rough estimate of its parameters?

Before proceeding, it is important to set up a suitable method to compare the solutions obtained by our approach DEP^{MEV} and the scenario-based one DEP^{SB}. In fact, a direct comparison of their optimal solution values could be misleading since they follow two different objective functions and, in the case of DEP^{SB}, the scenario set considered only approximates the real distribution of the stochastic parameters. For this reason, in all the following tests, given any first-stage solution $\boldsymbol{x}$ obtained through any of the two methods, we always calculate the corresponding value of a function $\Phi(\boldsymbol{x})$ defined as

$$\Phi(\boldsymbol{x}) \doteq -\sum_{i\in\mathcal{I}} c_i x_i + \frac{1}{|\bar{S}|}\sum_{s\in\bar{S}}\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{L}_i} x_i \cdot \max_{j\in\mathcal{L}_i}(V_{ij}+\theta^s_{ij}), \tag{32}$$

where $\bar{S}$ is a large set of scenarios such that $\bar{S}\cap S = \emptyset$ and $|\bar{S}| \gg |S|$. In particular, we set $|\bar{S}| = 10000$. The operator Φ computes the so-called *Sample Average Approximation* (SAA) performance of a first-stage solution $\boldsymbol{x}$ against possible future scenarios in $\bar{S}$ for our showcase problem (8), i.e., it assigns a uniform probability of $\frac{1}{|\bar{S}|}$ to each scenario and calculates its expected overall cost. It is important to note that the second-stage cost can always be obtained through the closed-form expression reported since, given a first-stage solution $\boldsymbol{x}$, the second-stage optimal solution for each scenario $s \in \bar{S}$ and each DCP $i \in \mathcal{I}$ corresponds to setting $y^s_{ij} = 1$ for the alternative $j \in \mathcal{L}_i$ that maximizes $V_{ij}+\theta^s_{ij}$ (all θ^s_{ij} are indeed known values at the second stage). Finally, the use of a large set of different scenarios (i.e., an *out-of-sample* perspective) mitigates potential overfitting in the creation of DEP^{SB} solutions but, more importantly, allows us to interpret the value of $\Phi(\boldsymbol{x})$ as the value of $\boldsymbol{x}$ for the *true* stochastic problem, which serves as a benchmark for the two approaches.

5.3.1. SP indicators

First, we want to assess how much the consideration of uncertainty affects a tsSP-DCP by computing classical SP measures such as the Expected Value of Perfect Information ($EVPI$) and the Value of the Stochastic Solution (VSS). The $EVPI$ [48] represents the value, averaged over all the future scenarios, obtainable if one could completely forecast the realization of the uncertainties. The VSS [49] represents instead the potential value improvement obtainable by optimizing the stochastic model instead of a deterministic one in which stochastic variables are substituted by their mean values. More precisely, we calculate such indicators as the percentages

$$EVPI[\%] = 100 \cdot \frac{WS - \Phi\left(\boldsymbol{x}^{MEV}\right)}{WS}$$

and

$$VSS[\%] = 100 \cdot \frac{\Phi\left(\boldsymbol{x}^{MEV}\right) - \Phi\left(\boldsymbol{x}^{EV}\right)}{\Phi\left(\boldsymbol{x}^{MEV}\right)},$$

where WS is the value of the so-called *wait-and-see* solution, i.e., the mean of the DEP^{SB} solved separately for each scenario, and $\boldsymbol{x}^{EV}$ is the optimal solution of the so-called *expected value* problem (EV), i.e., the problem (8) in which all the random variables are substituted by their mean values. Note that, usually, both the indicators are computed by considering $\boldsymbol{x}^{SB}$ as the benchmark solution. In contrast, in our case, the benchmark is $\boldsymbol{x}^{MEV}$ since Theorem 1 states it represents the optimal solution of model (8).

Figs. 1 and 2 report the boxplots describing, for different values of p and increasing values of ν, the $EVPI[\%]$ results for the L and NL instances, respectively. The $EVPI[\%]$ steadily increases as ν increases, while it decreases as p increases. Both behaviors are expected. In fact, the higher the value of ν, the higher the willingness to dispose of a perfect forecast of future realizations. Moreover, the overall uncertainty

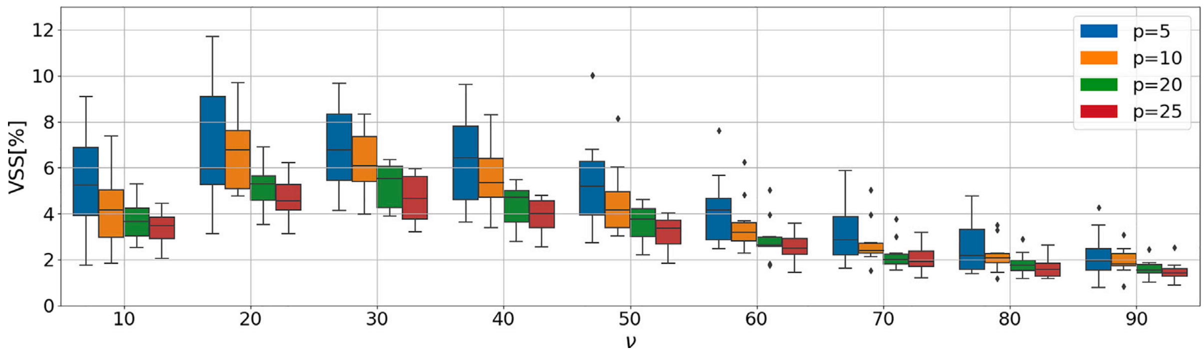

Fig. 3. $VSS[\%]$ against different values of p and ν for L instances.

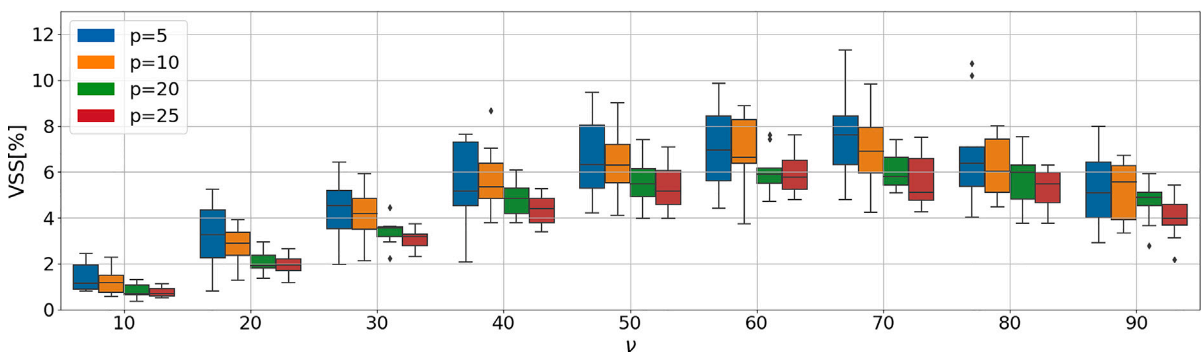

Fig. 4. $VSS[\%]$ against different values of p and ν for NL instances.

affecting the problem decreases when p increases, since a higher availability of facilities tends to average the impact of the stochasticity and reduces the possible opening combinations. In general, the $EVPI[\%]$ values range from 10% to more than 50% for L instances and from 2% to about 60% for NL ones. Actually, the results are quite close for the L and NL instances, with the former characterized by a slightly more consistent variance among the different runs (in particular in the case of small to medium values of ν). Moreover, while the growth of the $EVPI[\%]$ for the NL instances is almost linear with respect to the ν increasing, it has a logarithmic behavior in the case of L instances, reaching a stable value around $\nu = 50$.

Figs. 3 and 4 report the boxplots describing, for different values of p and increasing values of ν, the $VSS[\%]$ results for the L and NL instances, respectively. Similarly to the $EVPI[\%]$, the average $VSS[\%]$ decreases when p increases. Here, the size of the boxplots over the different runs allows us to appreciate that the variance of the results also tends to decrease when p increases (especially for the L instances). Again, by opening more facilities, it is possible to reduce the difference in terms of location choices between the stochastic problem and the EV. In general, however, the worst $VSS[\%]$ settles around 1%–2% while the highest values reach 10%–12%. On the other hand, there are some notable differences between the behavior of $VSS[\%]$ and that of the $EVPI[\%]$ when ν increases. In particular, for L instances, the average $VSS[\%]$ is around 4%–5% for the smallest value of ν and reaches a peak immediately when $\nu = 20$. After that, values steadily decrease to a 2% value when $\nu = 90$. On the contrary, for the NL instances, the smallest $VSS[\%]$ is obtained for $\nu = 10$, and the peak is obtained at $\nu = 70$. After that, values slightly decrease and settle at 4%–5% when $\nu = 90$.

The inverted U-shaped behavior of the $VSS[\%]$, differently from the monotonic one of the $EVPI[\%]$, can be explained by considering how parameter ν affects the tsSP-DCP structure and its optimal solution. For small values of ν, i.e., when the uncertainty is low, the EV solution $\mathbf{x}^{EV}$ well approximates the stochastic one $\mathbf{x}^{MEV}$, resulting in a low $VSS[\%]$. Then, as for any SP problem, by increasing the value of ν and thus the problem uncertainty, the two optimal solutions ($\mathbf{x}^{MEV}$ and $\mathbf{x}^{EV}$) show more evidently their peculiarities, resulting in more different optimal values and thus in a higher $VSS[\%]$. However, a specific characteristic of tsSP-DCPs is that, starting from a certain value of ν onward, the $VSS[\%]$ steadily decreases. In fact, as ν tends to its maximum value, the deterministic part of the utility of all the alternatives of all the DCPs becomes a negligible value, i.e., for $\nu \to 100$ we have $V_{ij} \to 0, \forall i \in \mathcal{I}, j \in \mathcal{L}_i$. This means that, for very large values of ν, all the second-stage DCPs tend to yield very similar utilities and, therefore, the corresponding first-stage decisions tend to become less important, since the second-stage recourse does not really depend on them. So, again, in this situation, the $\mathbf{x}^{EV}$ solution is expected to perform similarly to any other tsSP-DCP solution (including the optimal one $\mathbf{x}^{MEV}$), thus lowering the $VSS[\%]$. Clearly, the value of ν for which the concavity of the $VSS[\%]$ trend changes depends on the specific probability distribution used. From our results, we can see that this concavity change happens quite soon ($\nu \approx 20$) for the logit instances, while it moves onward ($\nu \approx 70$) for the nested logit ones, where the variance of the corresponding distribution is lower on average due to the presence of correlations between the alternatives within the same nest.

In conclusion, we experimentally proved that the simulated problems have a structure of uncertainty that justifies a stochastic approach and that there is enough room for improvement with respect to the EV for testing the effectiveness of our MEV-based approach, which is done in the following.

5.3.2. Quality and CPU performances of the proposed approach

We now compare the computational performance and the quality of the solutions found by solving model DEP^{MEV} against the solution of the model DEP^{SB} for different values of p.

First, let us define $\mathbf{x}^{SB}_{|S|}$ as the optimal solution of model (28) obtained by considering $|S|$ scenarios. Then, similarly as done above for the VSS and the $EVPI$, we compare the performance of $\mathbf{x}^{SB}_{|S|}$ and

Table 1
Percentage optimality gaps averaged over all the L instances.

	$\|S\|=10$		$\|S\|=50$		$\|S\|=100$		$\|S\|=500$		$\|S\|=1000$	
p	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$
5	3.05 (1.77)	0.83	2.01 (1.23)	0.83	1.51 (1.00)	0.83	0.83 (0.49)	1.62	0.56 (0.37)	1.67
10	2.65 (1.42)	0.85	1.63 (0.86)	1.67	1.21 (0.69)	0.84	0.74 (0.44)	1.67	0.51 (0.28)	3.33
20	1.99 (1.13)	0.00	1.29 (0.67)	2.50	0.98 (0.46)	1.69	0.62 (0.32)	2.50	0.42 (0.23)	5.00
25	1.84 (1.02)	1.67	1.18 (0.56)	1.67	0.91 (0.42)	1.64	0.56 (0.30)	5.83	0.37 (0.21)	4.17

Table 2
Percentage optimality gaps averaged over all the NL instances.

	$\|S\|=10$		$\|S\|=50$		$\|S\|=100$		$\|S\|=500$		$\|S\|=1000$	
p	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$	$Gap[\%]$	$n[\%]$
5	4.20 (2.69)	0.00	2.03 (1.71)	0.00	1.45 (1.64)	5.00	0.74 (0.90)	11.11	0.50 (0.62)	13.33
10	3.89 (2.33)	0.00	1.98 (1.64)	3.33	1.25 (1.18)	6.67	0.51 (0.57)	15.56	0.37 (0.44)	22.22
20	3.02 (1.88)	0.00	1.57 (1.31)	1.11	1.03 (0.92)	7.54	0.34 (0.35)	22.22	0.21 (0.25)	26.67
25	2.75 (1.77)	1.11	1.47 (1.23)	5.56	0.95 (0.89)	7.78	0.29 (0.33)	25.56	0.19 (0.23)	35.56

of $\boldsymbol{x}^{MEV}$ using the operator $\boldsymbol{\Phi}$, which uses a set of scenarios $\bar{S}$ much larger than S and which intersection with S must be empty. In fact, if we evaluated $\boldsymbol{\Phi}\left(\boldsymbol{x}^{SB}_{|S|}\right)$ on the same set of scenarios used to solve the DEP^{SB}, then $\boldsymbol{x}^{SB}_{|S|}$ would be unquestionably the optimal solution. Instead, operator $\boldsymbol{\Phi}$ allows us to evaluate the quality of the DEP^{SB} approximation of the stochastic problem (8), which is obtained by discretizing the probability distribution of θ_{ij}. In particular, we evaluate a percentage optimality gap calculated as

$$Gap[\%] \doteq 100 \cdot \frac{\boldsymbol{\Phi}\left(\boldsymbol{x}^{MEV}\right) - \boldsymbol{\Phi}\left(\boldsymbol{x}^{SB}_{|S|}\right)}{\boldsymbol{\Phi}\left(\boldsymbol{x}^{MEV}\right)}.$$

Since the sample average calculated by $\boldsymbol{\Phi}$ in (32) for assessing the optimal solution of model (28) is an unbiased estimator of the real average, we expect that the optimal solution value of DEP^{SB} is going to converge to the optimal solution of model DEP^{MEV} as the number of scenarios increases. We report the $Gap[\%]$ results obtained for L and NL instances in Table 1 and 2, respectively. In particular, for a specific combination of $|S|$ and p, each table presents the average $Gap[\%]$ and its standard deviation (in brackets) over all the L or NL instances, together with the percentage of times in which DEP^{MEV} and DEP^{SB} have a zero gap ($n[\%]$). We tested an increasing number of scenarios equal to 10, 50, 100, 500, and 1000.

First, it is interesting to highlight the very small percentage of times in which the two methods perform exactly in the same way, i.e., when the optimality gap is null. We can clearly see that this happens very rarely when considering up to 100 scenarios ($n[\%]$ never exceeds 2.5 for L instances and 7.8 for NL instances). When considering 500 or 1000 scenarios, $n[\%]$ values tend to increase; however, they never exceed 6% and 35% for L and NL instances, respectively. In general, null gaps are more consistent as p increases. Another clear trend we can derive is that the average and the standard deviation of the optimality gap both decrease as the value of p increases. Unsurprisingly, this also happens when the number of scenarios considered increases. With lower values of p, the gap is bigger since uncertainty has a greater relative impact on the solution when a small number of facilities must be selected to be located. When the number of scenarios is relatively small, optimality gaps can reach more than 4% on average, with less impact for the L instances, which seem easier to approximate. Interestingly enough, even considering 1000 scenarios, some instances with small values of p present a gap above 0.5%, and above 0.8% when considering 500 scenarios. This means that our method can be particularly useful when the number of facilities to locate is a small percentage of the total.

Finally, to better appreciate the $Gap[\%]$ trend and variance, we report in Figs. 5 and 6 boxplots of the results for the L and the NL instances, respectively. Each of the charts relates to a specific value of v, and data are separated for the value of the facility to locate and the number of considered scenarios. The values of v are chosen such that they correspond to the three notable situations of low, medium, or high VSS. Note that the v values differ in the two classes of instances, according to the results shown in Figs. 3 and 4.

The trends shown represent empirical evidence that the value of the optimal solution of DEP^{SB} tends to converge to the value of the optimal solution of DEP^{MEV} as $|S|$ increases. This is true for both L and NL instances, even if the convergence has different behaviors in the two classes and for the different values of v. Regarding the L instances, the convergence trends are quite smooth and regular, also concerning variances and differences against values of p. For $v = 10$, the convergence is so slow that the gaps settle at 2% even when using 1000 scenarios. For higher values of v, the convergence is faster; however, in both cases, DEP^{SB} obtains optimality gaps around 1% only with 500 scenarios and around 0.5% with 1000 scenarios. Using 100 scenarios leads to gaps up to 3% for any value of v and p, while using 50 scenarios leads to gaps up to 5%. Note also that the fewer the scenarios used, the higher the variability of the gaps and, thus, the lower the trustworthiness of the corresponding solutions. We remark that obtaining a reasonable degree of solution quality in a reasonable time using 500 scenarios for many stochastic combinatorial optimization problems could be a challenging task against real-life instance dimensions, sometimes totally out of reach. If we look at NL instances, the results are more erratic both concerning averages and variances. For $v = 20$, the convergence is quite fast, and gaps rarely exceed 2%. Instead, for higher values of v, the optimality gaps are substantial (up to 9%) if we consider 100 scenarios or fewer. For $v = 90$, even with 1000 scenarios, the average gaps never settle below 1%. Finally, all these trends confirm that the effectiveness of our approach is emphasized when the number of facilities to be located tends to be small.

Finally, to assess the computational performances of our method, we compute the percentage loss in CPU time to solve DEP^{SB} instead of DEP^{MEV} as

$$Gap\ Time[\%] \doteq 100 \cdot \frac{t^{SB}_{|S|} - t^{MEV}}{t^{SB}_{|S|}}, \tag{33}$$

where $t^{SB}_{|S|}$ is the computational time required to solve DEP^{SB} considering $|S|$ scenarios and t^{MEV} is the one required to solve DEP^{MEV}. The relative results are reported in Figs. 7 and 8 for the L instances and the NL instances, respectively, and each chart corresponds to the notable v values already described. Instances are disaggregated per the number of scenarios and the number of facilities to locate. Due to the difference in magnitude of the values when the number of scenarios increases, the y axis is in logarithmic scale.

The results for both the instance classes and for all the values of v are really close. The main evidence is, obviously, the enormous increase in the CPU time gap as the number of scenarios considered when solving the DEP^{SB} increases. When $|S| = 10$, DEP^{SB} requires on

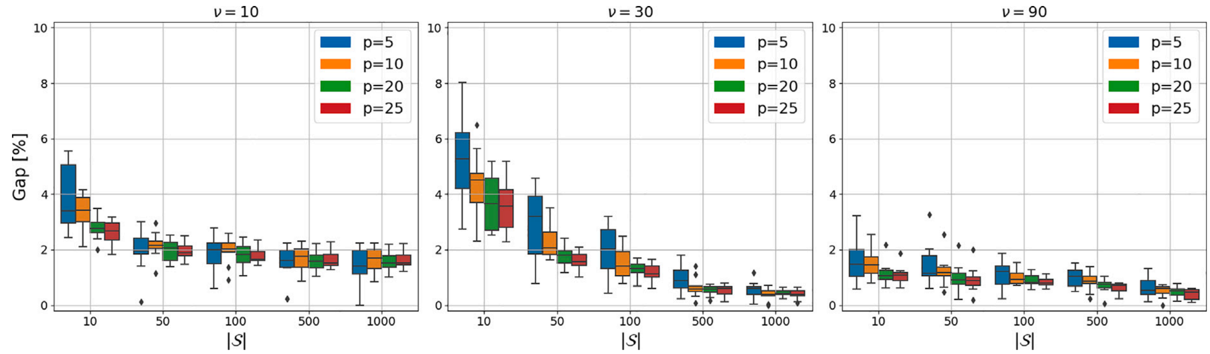

Fig. 5. Percentage optimality gaps for L instances.

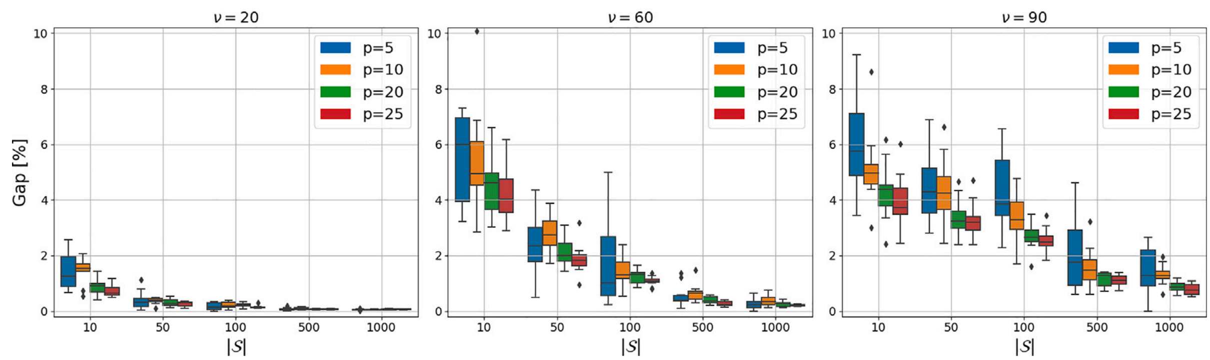

Fig. 6. Percentage optimality gaps for NL instances.

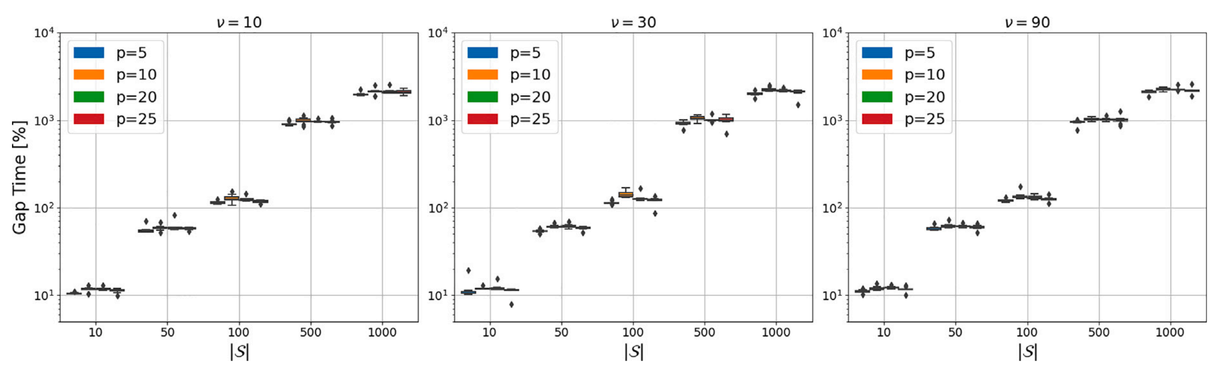

Fig. 7. Computational time gaps for L instances.

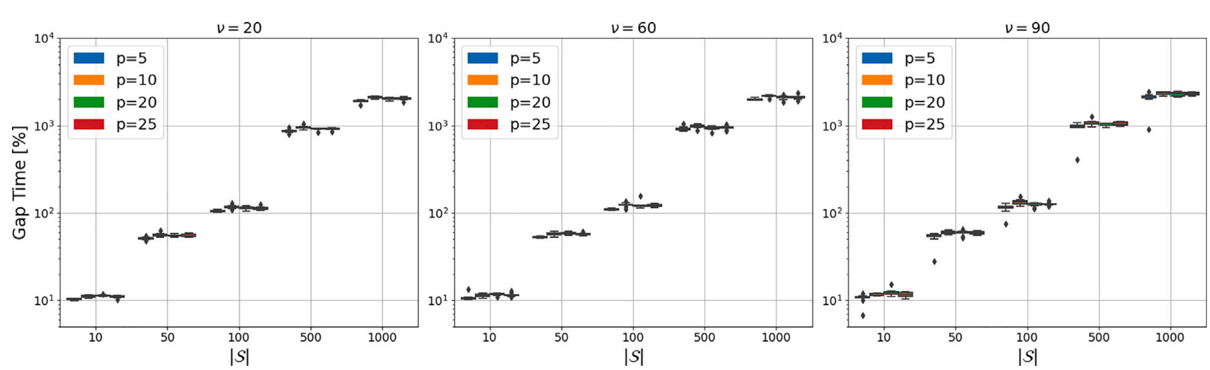

Fig. 8. Computational time gaps for NL instances.

average more than 10 times the computational time of DEP^{MEV}, and this value increases up to 2000 times when $|S| = 1000$. A particularly great difference is shown in all the cases when moving from 100 to 500 scenarios, with times that are basically one order of magnitude greater. The incredible efficiency of our method is because, as already stated, the DEP^{SB} formulation is strongly affected by the number of scenarios

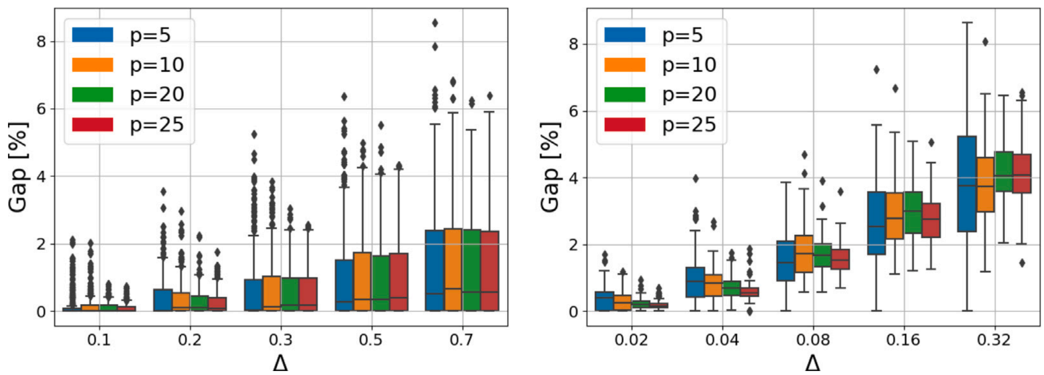

Fig. 9. Optimality gaps against calibration errors on β for L instances (left chart) and on all $\beta_m, m = 1, \dots, r$ for NL instances (right chart).

$|S|$ in terms of the number of variables and constraints, while DEP^{MEV} is not at all.

5.3.3. Robustness of the framework against distributions' parameters calibration

In all those practical cases in which we are not working with a pure theoretical probability distribution of the random variables, the θ_{ij} for each $i \in \mathcal{I}$ and each $j \in \mathcal{L}_i$ are distributed according to parameters that may be unknown and must be derived or calibrated from the available historical data. This may lead to off values of such parameters and, in turn, to solutions that are not necessarily optimal when using our deterministic equivalent form.

In the following, we want to understand how robust our framework is against possible wrong values of all the parameters needed to define the generating function of the MEV probability distribution used, given that all the other conditions still hold. Hence, given each known theoretical parameter λ^t on which stochastic data are generated for each tested instance, we solve the DEP^{MEV} by using different values of λ randomly drawn in the range $[\lambda^t - \Delta\lambda^t, \lambda^t + \Delta\lambda^t]$. Clearly, by varying Δ, we can control the percentage deviation of the λ actually used with respect to the theoretical one. In particular, in our experiments, L instances need a single parameter to be estimated, namely, the homogeneity parameter β, while NL instances need to calibrate r parameters, namely, a β_m value for each nest $m = 1, \dots, r$. Then, for all the instances, we calculate the percentage optimality gap

$$Gap[\%] \doteq 100 \cdot \frac{\left|\Phi\left(\mathbf{x}^{MEV(\lambda^t)}\right) - \Phi\left(\mathbf{x}^{MEV(\lambda)}\right)\right|}{\Phi\left(\mathbf{x}^{MEV(\lambda^t)}\right)}$$

where $\mathbf{x}^{MEV(\lambda)}$ represents the optimal solution of DEP^{MEV} using the randomly drawn vector of parameters λ while $\mathbf{x}^{MEV(\lambda^t)}$ represents that using the vector of theoretical parameters λ^t. Note that in the $Gap[\%]$ formula, the difference in the numerator is taken in absolute value since no properties ensure that the obtained empirical solution represents an upper or a lower bound for the optimal theoretical one. The boxplot results of $Gap[\%]$ for the L and NL instances, against increasing values of Δ and separated per p values, are reported in Fig. 9.

Concerning L instances, we can see that our framework provides very robust solutions. In fact, up to 20% error on theoretical β, the optimality gaps are negligible, with averages very little above 0 and very few outliers above 2%. Even increasing the value of Δ up to 70%, the average gaps always stay below 1%, with some worst values about 6%. Interestingly enough, the value of p does not affect the robustness of the framework, with basically no differences between the relative boxplots (only the dispersion of the outliers seems to decrease with the increase of p). The results for NL instances are slightly more erratic, even if the average gaps do not exceed 4% against parameters that are off by up to more than 30% of their values. Clearly, the robustness variance for the NL instances strongly depends on the difficulty of calibrating more than one parameter (one for each nest) and on the different propagation of the calibration error due to alternative correlations.

A significant conclusion of the above robustness analysis is that, in many cases and up to a consistent percentage of errors on all the parameters to estimate, the gaps settle below the relative values obtainable by the scenario-based method and using a reasonable number of scenarios. This can be seen by comparing the percentage gaps of Fig. 9 and those reported in Figs. 5 and 6. In particular, it seems that on average, even with a strongly off calibration of β on L instances (up to 30%), the effectiveness of our approach is maintained against using scenario-based simulations with up to 100 scenarios. Concerning NL instances, up to a calibration error of 10%, our method seems much more effective than the scenario-based one using up to 100 scenarios in the case of medium to large values of ν. Such results confirm the usefulness of our approach in all practical cases in which the real probability distribution of the random variables is not known or known only by a small part.

6. Beyond tsSP-DCPs: the tsSP-DEV *template* model and its deterministic equivalent form

Before concluding, we want to show the potentiality of applying the theoretical results presented in this paper beyond the class of tsSP-DCPs. More precisely, we define hereafter a larger class of two-stage Stochastic Programming problems (tsSP) whose expected second-stage optimum is decomposable (D) into a finite number of expectations of extreme values (EVs), which we name tsSP-DEVs. We will show that a MEV-based deterministic equivalent form can be obtained for a tsSP-DEV too, although accepting an increase in its computational complexity.

Let us consider n first-stage decision variables $x_i, i = 1, \dots, n$, and n sequences of random variables, each of length $l_i, i = 1, \dots, n$. Within the ith sequence, each random variable in position $j = 1, \dots, l_i$ is

$$U_{ij}(x_i, \theta_{ij}) \doteq V_{ij}(x_i) + \theta_{ij},$$

where $V_{ij}(x_i)$ is a deterministic utility function and θ_{ij} is a stochastic oscillation. Note that, differently from a tsSP-DCP, here the deterministic utility may depend on the first-stage decision variables' values, thus embracing more general utility functions. A two-stage SP problem as defined in (1) is a tsSP-DEV when the second-stage expected value $\mathbb{E}_{\theta}[h(\mathbf{x}, \boldsymbol{\theta})]$ can be written as

$$\mathbb{E}_{\theta}[h(\mathbf{x}, \boldsymbol{\theta})] = g(\mathbf{x}, \mathbb{E}_{\theta}[H_1(x_1, \boldsymbol{\theta}_1)], \dots, \mathbb{E}_{\theta}[H_n(x_n, \boldsymbol{\theta}_n)]) \quad (34)$$

where $H_i(x_i, \boldsymbol{\theta}_i) = \max_{j=1,\dots,l_i} U_{ij}(x_i, \theta_{ij})$ is the *extreme value* of the ith sequence of random variables, $\boldsymbol{\theta}_i$ is the vector of random variables $[\theta_{i,1}, \dots, \theta_{i,l_i}]^T$, and g is any scalar function. The fundamental characteristic of a tsSP-DEV is that its second stage can be decomposed into n stochastic optimization problems, the optimum of which can be stated as the maximum of a sequence of random variables whose realizations depend on the first-stage decisions. The function g can be seen as a *pooling* strategy applied to the expected optima deriving from the resolution of the n problems in order to obtain the total second-stage expected optimum.

Under the same assumptions on the probability distribution of the stochastic parameters that allow to capture the decision-maker choice behavior through an MEV model (see Section 4.1), the special structure of a tsSP-DEV makes it possible to derive a deterministic equivalent form, i.e.,

$$\max_{\boldsymbol{x}} \quad f_1(\boldsymbol{x}) + g(\boldsymbol{x}, \mathbb{E}_\theta[H_1(x_1,\boldsymbol{\theta}_1)], \dots, \mathbb{E}_\theta[H_n(x_n,\boldsymbol{\theta}_n)])$$
$$\text{s.t.} \quad \boldsymbol{x} \in \mathcal{X}$$

where

$$\mathbb{E}_\theta[H_i(x_i,\boldsymbol{\theta}_i)] = \frac{1}{\beta}\left(\ln\left(G_i\left(e^{V_{i,1}(x_i)}, \dots, e^{V_{i,m_i}(x_i)}\right)\right) + \gamma\right), \ \forall i = 1,\dots,n.$$

This can be seen by applying trivial modifications to Theorem 2 and its proof.

It is important to notice that the just-defined tsSP-DEV framework is a sort of *template* model from which it is possible to derive factual classes of stochastic problems of interest only by analytically defining the deterministic utility function $V_{ij}(x_i), \forall j = 1,\dots,l_i, \forall i = 1,\dots,n$, and the pooling function g. It is easy to see that a tsSP-DCP represents a tsSP-DEV special case, since the two mentioned degrees of freedom have been fixed. More precisely, the second-stage expected optimum of a tsSP-DCP derived in (22) can be written in the form (34) where g is the linear combination

$$g(\boldsymbol{x}, \mathbb{E}_\theta[H_1(x_1,\boldsymbol{\theta}_1)], \dots, \mathbb{E}_\theta[H_n(x_n,\boldsymbol{\theta}_n)]) = \left[\mathbb{E}_\theta[H_1(x_1,\boldsymbol{\theta}_1)], \dots, \mathbb{E}_\theta[H_n(x_n,\boldsymbol{\theta}_n)]\right]^T \cdot \boldsymbol{x}$$

and where $H_i(x_i,\theta_i) = H_i(\theta_i), \forall i = 1,\dots,n$, since the deterministic utilities are constant, i.e.,

$$V_{ij}(x_i) = V_{ij}, \forall j = 1,\dots,l_i, \forall i = 1,\dots,n.$$

However, a broad interpretation of tsSP-DEV models can be inferred from their general properties. In many SP applications, the decision-maker is interested in finding a strategy that maximizes the expected total profit (or minimizes the expected total cost), and, therefore, the second-stage optimum is defined as the maximum (or the minimum) of a function involving the summation of several expressions that depend on random variables. A tsSP-DEV, instead, focuses on optimizing a function of the expectations of extreme situations (i.e., the maxima or minima of sequences of random variables) rather than expected behaviors. In a sense, and similar to the problems addressed in Bali [50], Bertsimas et al. [51], or Bertsimas et al. [52], a tsSP-DEV involves characteristics both from the classical SP paradigm and from the Robust Optimization one. In fact, the tsSP-DEV optimization perspective is still focused on making here-and-now decisions by also evaluating the expected value of an uncertain problem over all the possible future situations, but assuming that such a value can be obtained as a function of extreme future behaviors. Hence, tsSP-DEVs can be used for applications such as those related to the modeling of certain natural catastrophes, the management of financial risks, and the modeling of rational choices (as in the tsSP-DCP case).

As a final aspect, we want to highlight that the computational complexity of the deterministic equivalent program obtained through our approach strongly depends on the structure of the function g and on the way such a function and the deterministic utilities V_{ij} depend on the first-stage variables. In general, even if the deterministic part of the utilities linearly depends on the first-stage decision variables (as for tsSP-DCPs), the resulting deterministic equivalent problem may be highly non-linear, and thus possibly very hard to solve. On the other hand, the approach still allows for totally getting rid of the complexity generated by the problem uncertainty, thus enabling the development of ad hoc deterministic exact or approximation methods.

7. Conclusions

In this work, we closely addressed a class of stochastic problems called two-stage Stochastic Programs embedding Discrete Choice Problems (tsSP-DCPs), in which independent DCPs appear as the second stage and are triggered and/or weighted by first-stage decisions. Despite their simple combinatorial structure, tsSP-DCPs can be found within many managerial settings and applications. Under the typical assumptions needed to obtain MEV choice models, we have been able to derive a closed-form analytical expression for the expected value of the second stage and, in turn, to obtain a deterministic problem equivalent to any tsSP-DCP. An extensive experimental analysis of a showcase problem has validated our approach both in terms of precision and efficiency against classical scenario-based deterministic approximations. In particular, we have shown that the more the decision-maker wants to accurately represent the uncertainty through scenarios, the more our deterministic formulation represents a faster alternative. The computational improvement is so high that there is even room for evaluating approximate versions of our approach. Finally, a discussion of the approach's applicability and generality highlighted its possible use when little knowledge is available on uncertain parameters and its potential to address more complex stochastic optimization problems.

Future works regard the application of the approach (or its approximate version) to tsSP-DCPs appearing as the main setting or as a subproblem in specific practical contexts. The area of Social Engagement optimization [53,54], in which customers' choices are involved, seems of particular interest. In this regard, in the presence of additional complicating constraints, they can be relaxed through Lagrangian or other techniques to comply with the tsSP-DCP assumptions, while, against deterministic equivalent problems still very complex to solve, specific methodologies collaborating with our analytical approach can be developed. Moreover, the generalization of our method to an operational multi-stage SP context (i.e., having several stages much greater than two) could be addressed more specifically. Finally, the tsSP-DEV class of problems could be further explored, and interesting properties could be found between ad-hoc generating functions and the relative stochastic optimization problem (as they exist in choice modeling for the logit-like family of models).

CRediT authorship contribution statement

Michel Bierlaire: Writing – review & editing, Supervision, Methodology, Conceptualization. **Edoardo Fadda:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Lohic Fotio Tiotsop:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Daniele Manerba:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The last author gratefully acknowledges the support of “ULTRA OPTYMAL - Urban Logistics and sustainable TRAnsportation: OPtimization under uncertainTY and MAchine Learning”, a PRIN 2020 project funded by the Italian University and Research Ministry (number 20207C8T9M, website: https://ultraoptymal.unibg.it). All the authors dedicate this paper to the memory of Roberto Tadei, Emeritus Professor at Politecnico di Torino (Italy), who concretely enabled and supported the collaboration that has led to this research work.

Appendix. Expected maximum utility in MEV models

The following theorem allows us to calculate the expected maximum utility in MEV models. This result rests on the Corollary of Theorem 1 derived in McFadden [13]. The statement and the proof have been reformulated in line with our decisional setting and the same notation used in our paper.

Theorem 2. *Consider a DCP i, as formulated in (3), in which the decision-maker's choice behavior is described by an MEV model, i.e., there exists a β–homogeneous generating function G_i ensuring Assumptions 1 and 2. Then, it holds:*

$$\mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] = \frac{1}{\beta}\left(\ln\left(G_i(e^{V_{i,1}},\dots,e^{V_{i,|\mathcal{L}_i|}})\right)+\gamma\right), \tag{A.1}$$

where $\gamma = -\int_0^{+\infty}\ln(t)e^{-t}dt$ is the Euler constant.

Proof. Let us define E_j as the event in which j is the alternative with the maximum utility within its set, i.e., $j = \operatorname{argmax}_{k\in\mathcal{L}_i} U_{ik}(\theta_{ik})$. Then, the expectation we are looking for can be written as conditioned by the partition of the sample space described by E_j, i.e.:

$$\mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] = \sum_{j\in\mathcal{L}_i}\mathbb{E}_\theta\left[U_{ij}(\theta_{ij})|E_j\right]\cdot\mathbb{P}\left[E_j\right]. \tag{A.2}$$

First, observe that, given a $\bar{j}\in\mathcal{L}_i$, for each $j\in\mathcal{L}_i$ it holds

$$U_{ij}(\theta_{ij}) \ge U_{i\bar{j}}(\theta_{i\bar{j}}) \iff V_{ij}+\theta_{ij} \ge V_{i\bar{j}}+\theta_{i\bar{j}} \iff V_{ij}-V_{i\bar{j}}+\theta_{ij}\ge\theta_{i\bar{j}}$$

and, in turn, the probability density function $f_{i|E_j}(u_j)$ of θ_{ij} given the event E_j is

$$\begin{aligned} f_{i|E_j}(u_j) &= \frac{1}{\mathbb{P}[E_j]}\cdot\\ &\cdot\int_{z_1=-\infty}^{V_{ij}-V_{i,1}+u_j}\cdots\int_{z_{j-1}=-\infty}^{V_{ij}-V_{i,j-1}+u_j}\int_{z_{j+1}=-\infty}^{V_{ij}-V_{i,j+1}+u_j}\cdots\int_{z_{|\mathcal{L}_i|}=-\infty}^{V_{ij}-V_{i,|\mathcal{L}_i|}+u_j} f_i(z_1,\dots,z_{|\mathcal{L}_i|})\,d\mathbf{z}_{-j} =\\ &= \frac{1}{\mathbb{P}[E_j]}\frac{\partial F_i}{\partial z_j}(V_{ij}-V_{i,1}+u_j,\dots,u_j,\dots,V_{ij}-V_{i,|\mathcal{L}_i|}+u_j) \end{aligned} \tag{A.3}$$

where $d\mathbf{z}_{-j} = dz_1\dots dz_{j-1}dz_{j+1}\dots dz_{|\mathcal{L}_i|}$ and $f_i(z_1,\dots,z_{|\mathcal{L}_i|})$ is the probability density function of $(\theta_{i,1},\dots,\theta_{i,|\mathcal{L}_i|})$. Using A.3, we have that

$$\begin{aligned} \mathbb{E}_\theta\left[U_{ij}(\theta_{ij})|E_j\right] &= \int_{u_j=-\infty}^{+\infty} U_{ij}(u_j)f_{i|E_j}(u_j)\,du_j =\\ &= \frac{1}{\mathbb{P}\left[E_j\right]}\int_{u_j=-\infty}^{+\infty} U_{ij}(u_j)\frac{\partial F_i}{\partial z_j}(V_{ij}-V_{i,1}+u_j,\dots,u_j,\dots,V_{ij}-V_{i,|\mathcal{L}_i|}+u_j)\,du_j. \end{aligned} \tag{A.4}$$

Now, by plugging A.4 in A.2, we get

$$\begin{aligned} \mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] &= \sum_{j\in\mathcal{L}_i}\mathbb{E}_\theta\left[U_{ij}(\theta_{ij})|E_j\right]\cdot\mathbb{P}\left[E_j\right] =\\ &= \sum_{j\in\mathcal{L}_i}\int_{u_j=-\infty}^{+\infty} U_{ij}(u_j)\frac{\partial F_i}{\partial z_j}(V_{ij}-V_{i,1}+u_j,\dots,u_j,\dots,V_{ij}-V_{i,|\mathcal{L}_i|}+u_j)\,du_j. \end{aligned} \tag{A.5}$$

Since we are dealing with multivariate extreme value distributions (see Assumption 1), the derivative in A.5 becomes

$$\begin{aligned} &\frac{\partial F_i}{\partial z_j}(V_{ij}-V_{i,1}+u_j,\dots,u_j,\dots,V_{ij}-V_{i,|\mathcal{L}_i|}+u_j) =\\ &= \frac{\partial \exp\left(-G_i\left(e^{-(V_{ij}-V_{i,1}+u_j)},\dots,e^{-u_j},\dots,e^{-(V_{ij}-V_{i,|\mathcal{L}_i|}+u_j)}\right)\right)}{\partial z_j} =\\ &= e^{-u_j}\frac{\partial G_i}{\partial z_j}\left(e^{-(V_{ij}-V_{i,1}+u_j)},\dots,e^{-u_j},\dots,e^{-(V_{ij}-V_{i,|\mathcal{L}_i|}+u_j)}\right)\cdot\\ &\quad\cdot\exp\left(-G_i\left(e^{-(V_{ij}-V_{i,1}+u_j)},\dots,e^{-u_j},\dots,e^{-(V_{ij}-V_{i,|\mathcal{L}_i|}+u_j)}\right)\right). \end{aligned} \tag{A.6}$$

Now, by replacing $u_j = V_{ij}-V_{ij}+u_j$, using Property 1 to extract the term $e^{-\beta(V_{ij}+u_j)}$ from G_i, and using Property 1 together with Property 3 to extract $e^{-(\beta-1)(V_{ij}+u_j)}$ from $\frac{\partial G_i}{\partial z_j}$, we get that A.6 is equivalent to

$$\begin{aligned} &e^{-u_j}e^{-(\beta-1)(V_{ij}+u_j)}\frac{\partial G_i}{\partial z_j}\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\cdot\\ &\qquad\cdot\exp\left(-e^{-\beta(V_{ij}+u_j)}G_i\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\right) =\\ &= e^{-\beta(V_{ij}+u_j)}e^{V_{ij}}\frac{\partial G_i}{\partial z_j}\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\cdot\\ &\qquad\cdot\exp\left(-e^{-\beta(V_{ij}+u_j)}G_i\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\right). \end{aligned}$$

This means that, for each $j\in\mathcal{L}_i$, the integral in A.5 becomes

$$\begin{aligned} &\int_{u_j=-\infty}^{+\infty}(V_{ij}+u_j)\frac{\partial F_i}{\partial z_j}(V_{ij}-V_{i,1}+u_j,\dots,u_j,\dots,V_{ij}-V_{i,|\mathcal{L}_i|}+u_j)\,du_j =\\ &= \int_{u_j=-\infty}^{+\infty}(V_{ij}+u_j)e^{-\beta(V_{ij}+u_j)}e^{V_{ij}}\frac{\partial G_i}{\partial z_j}\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\cdot\\ &\qquad\cdot\exp\left(-ae^{-\beta(V_{ij}+u_j)}\right)du_j, \end{aligned}$$

where we have defined $a \doteq G_i\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)$ for notational convenience.

By noting that $e^{V_{ij}}\frac{\partial G_i}{\partial z_j}\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)$ does not depend on u_j, A.5 becomes

$$\begin{aligned} &\mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] =\\ &= \sum_{j\in\mathcal{L}_i} e^{V_{ij}}\frac{\partial G_i}{\partial z_j}\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\cdot\\ &\qquad\cdot\int_{u_j=-\infty}^{+\infty}(V_{ij}+u_j)e^{-\beta(V_{ij}+u_j)}\exp(-ae^{-\beta(V_{ij}+u_j)})\,du_j. \end{aligned} \tag{A.7}$$

Since each function in the set $\{(V_{ij}+u_j)e^{-\beta(V_{ij}+u_j)}\exp(-ae^{-\beta(V_{ij}+u_j)})\}_{j\in\mathcal{L}_i}$ is an horizontal translation of every other, and since we are integrating over all $\mathbb{R}$, we can substitute $w = V_{ij}+u_j$, obtaining

$$\begin{aligned} &\mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] =\\ &= \sum_{j\in\mathcal{L}_i} e^{V_{ij}}\frac{\partial G_i}{\partial z_j}\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right)\int_{w=-\infty}^{+\infty} we^{-\beta w}\exp(-ae^{-\beta w})\,dw =\\ &= \left(\int_{w=-\infty}^{+\infty} we^{-\beta w}\exp(-ae^{-\beta w})\,dw\right)\sum_{j\in\mathcal{L}_i} e^{V_{ij}}\frac{\partial G_i}{\partial z_j}\left(e^{V_{i,1}},\dots,e^{V_{ij}},\dots,e^{V_{i,|\mathcal{L}_i|}}\right). \end{aligned}$$

Then, applying Euler's Theorem (Property 2) on the second factor, we get

$$\begin{aligned} \mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i} U_{ij}(\theta_{ij})\right] &= \left(\int_{w=-\infty}^{+\infty} we^{-\beta w}\exp(-ae^{-\beta w})\,dw\right)\beta a =\\ &= \int_{w=-\infty}^{+\infty} w\beta ae^{-\beta w}\exp(-ae^{-\beta w})\,dw. \end{aligned} \tag{A.8}$$

Now, let us make a variable change in A.8 by defining $t = ae^{-\beta w}$. Therefore, we have

$$w = \frac{\ln(a)-\ln(t)}{\beta}$$

and

$$dt = -\beta ae^{-\beta w}dw.$$

Since $w\to+\infty$ implies $t\to0$ and $w\to-\infty$ implies $t\to+\infty$, A.8 is equivalent to

$$-\frac{1}{\beta}\int_{t=+\infty}^{0}\left(\ln(a)-\ln(t)\right)e^{-t}\,dt = \frac{1}{\beta}\left(\ln(a)\int_{t=0}^{+\infty}e^{-t}dt - \int_{t=0}^{+\infty}\ln(t)e^{-t}\,dt\right). \tag{A.9}$$

Recalling that $-\int_0^{+\infty}\ln(t)e^{-t}dt$ is the Euler constant γ and $\int_0^{+\infty}e^{-t}dt = 1$, we finally get

$$\mathbb{E}_\theta\left[\max_{j\in\mathcal{L}_i}U_{ij}(\theta_{ij})\right] = \frac{1}{\beta}\left(\ln\left(G_i(e^{V_{i,1}},\ldots,e^{V_{ij}},\ldots,e^{V_{i,|\mathcal{L}_i|}})\right)+\gamma\right), \tag{A.10}$$

which proves the thesis. □

Data availability

Data will be made available on request.

References

[1] Wallace SW, Ziemba WT. Applications of stochastic programming. In: MPS-SIAM series on optimization, Society for Industrial and Applied Mathematics; 2005.
[2] King AJ, Wallace SW. Modeling with stochastic programming. In: Springer series in operations research and financial engineering, Springer-Verlag New York; 2012.
[3] Beraldi P, Bruni ME, Manerba D, Mansini R. A stochastic programming approach for the traveling purchaser problem. IMA J Manag Math 2017;28(1):41–63.
[4] Manerba D, Mansini R, Perboli G. The capacitated supplier selection problem with total quantity discount policy and activation costs under uncertainty. Int J Prod Econ 2018;198:119–32.
[5] Giusti R, Manerba D, Tadei R. Multiperiod transshipment location–allocation problem with flow synchronization under stochastic handling operations. Networks 2021;78(1):88–104.
[6] Li J, Deshpande A. Maximizing expected utility for stochastic combinatorial optimization problems. Math Oper Res 2019;44(1):354–75.
[7] Birge JR, Louveaux FV. Introduction to stochastic programming. New York: Springer Verlag; 1997.
[8] Ben-Akiva ME, Lerman SR. Discrete choice analysis: theory and application to travel demand. MIT Press; 1985.
[9] Wang R. Discrete choice models and applications in operations management. In: Tutorials in operations research: emerging optimization methods and modeling techniques with applications. INFORMS; 2021, p. 199–226.
[10] Tadei R, Perboli G, Ricciardi N, Baldi M. The capacitated transshipment location problem with stochastic handling utilities at the facilities. Int Trans Oper Res 2012;19:789–807.
[11] Perboli G, Tadei R, Baldi MM. The stochastic generalized bin packing problem. Discrete Appl Math 2012;160(7):1291–7.
[12] Fadda E, Fotio Tiotsop L, Manerba D, Tadei R. The stochastic multi-path traveling salesman problem with dependent random travel costs. Transp Sci 2020;54(5):1372–87.
[13] McFadden D. Modelling the choice of residential location. In: Karlquist A, et al., editors. Spatial interaction theory and residential location. Amsterdam: North-Holland; 1978, p. 75–96.
[14] Fosgerau M, McFadden D, Bierlaire M. Choice probability generating functions. J Choice Model 2013;8:1–18.
[15] Powell WB. A unified framework for stochastic optimization. European J Oper Res 2019;275(3):795–821.
[16] Zhou R, Bhuiyan TH, Medal HR, Sherwin MD, Yang D. A stochastic programming model with endogenous uncertainty for selecting supplier development programs to proactively mitigate supplier risk. Omega 2022;107:102542.
[17] Li Y, Liu M, da Gama FS, Yang Z. Risk-averse two-stage stochastic programming for assembly line reconfiguration with dynamic lot sizes. Omega 2024;127:103092.
[18] Ji H, Wang R, Zhang C, Yin J, Ma L, Yang L. Optimization of train schedule with uncertain maintenance plans in high-speed railways: A stochastic programming approach. Omega 2024;124:102999.
[19] Ntaimo L, Sen S. The million-variable "march" for stochastic combinatorial optimization. J Global Optim 2005;32(3):385–400.
[20] Petruzzi NC, Dada M. Pricing and the newsvendor problem: A review with extensions. Oper Res 1999;47(2):183–94.
[21] Bianchi L, Dorigo M, Gambardella LM, Gutjahr WJ. A survey on metaheuristics for stochastic combinatorial optimization. Nat Comput 2009;8(2):239–87.
[22] Hua G, Zhang Y, Cheng T, Wang S, Zhang J. The newsvendor problem with barter exchange. Omega 2020;92:102149.
[23] McFadden D. Conditional logit analysis of qualitative choice behavior. Front Econ 1973;105–42.
[24] Hoetker G. The use of logit and probit models in strategic management research: Critical issues. Strat Manag J 2007;28(4):331–43.
[25] Chen X, Wang Y, Zhou Y. Optimal policy for dynamic assortment planning under multinomial logit models. Math Oper Res 2021;46(4):1639–57.
[26] Gallego G, Wang R. Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. Oper Res 2014;62(2):450–61.
[27] Du C, Cooper WL, Wang Z. Optimal pricing for a multinomial logit choice model with network effects. Oper Res 2016;64(2):441–55.
[28] Manchiraju C, Dawande M, Janakiraman G. Multiproduct pricing with discrete price sets. Oper Res 2022;70(4):2185–93.
[29] Chen M. Stochastic optimization and applications with endogenous uncertainties via discrete choice models Electronic Theses and Dissertations, (6741). 2019, URL: https://stars.library.ucf.edu/etd/6741.
[30] Pacheco Paneque M, Bierlaire M, Gendron B, Azadeh SS. Integrating advanced discrete choice models in mixed integer linear optimization. Transp Res Part B: Methodol 2021;146:26–49.
[31] Bortolomiol S, Lurkin V, Bierlaire M. A simulation-based heuristic to find approximate equilibria with disaggregate demand models. Transp Sci 2021;55(5):1025–45.
[32] Bortolomiol S, Lurkin V, Bierlaire M. Price-based regulation of oligopolistic markets under discrete choice models of demand. Transportation 2022;49(5):1441–63.
[33] Galambos J. The asymptotic theory of extreme order statistics. Wiley series in probability and mathematical statistics, R.E. Krieger Publishing Company; 1987.
[34] Galambos J, Lechner J, Simiu E. Extreme value theory and applications. Springer New York, NY; 1994.
[35] Tadei R, Ricciardi N, Perboli G. The stochastic p-median problem with unknown cost probability distribution. Oper Res Lett 2009;37(2):135–41.
[36] Perboli G, Tadei R, Gobbato L. The multi-handler knapsack problem under uncertainty. European J Oper Res 2014;236(3):1000–7.
[37] Tadei R, Perboli G, Perfetti F. The multi-path traveling salesman problem with stochastic travel costs. EURO J Transp Logist 2017;6(1):3–23.
[38] Tadei R, Perboli G, Manerba D. The multi-stage dynamic stochastic decision process with unknown distribution of the random utilities. Optim Lett 2020;14:1207–18.
[39] Roohnavazfar M, Manerba D, De Martin JC, Tadei R. Optimal paths in multi-stage stochastic decision networks. Oper Res Perspect 2019;6:100124.
[40] Roohnavazfar M, Manerba D, Fotio Tiotsop L, Pasandideh SHR, Tadei R. Stochastic single machine scheduling problem as a multi-stage dynamic random decision process. Comput Manag Sci 2021;18:267–97.
[41] Daly A, Zachary S. Improved multiple choice models. In: Hensher D, Dalvi Q, editors. Determinants of travel choice. Sussex: Saxon House; 1978.
[42] Mishra VK, Natarajan K, Padmanabhan D, Teo C-P, Li X. On theoretical and empirical aspects of marginal distribution choice models. Manag Sci 2014;60(6):1511–31.
[43] Zhang H, Rusmevichientong P, Topaloglu H. Technical note—Multiproduct pricing under the generalized extreme value models with homogeneous price sensitivity parameters. Oper Res 2018;66(6):1559–70.
[44] Guevara CA, Ben-Akiva ME. Sampling of alternatives in Multivariate Extreme Value (MEV) models. Transp Res Part B: Methodol 2013;48:31–52.
[45] Bierlaire M. A theoretical analysis of the cross-nested logit model. Ann Oper Res 2006;144(1):287–300.
[46] Tadei R, Perboli G, Manerba D. A recent approach to derive the multinomial logit model for choice probability. AIRO Springer Ser 2018;1:473–81.
[47] Stephenson A. Simulating multivariate extreme value distributions of logistic type. Extremes 2003;6:49–59.
[48] Dempster MAH. The expected value of perfect information in the optimal evolution of stochastic systems. In: Arató M, Vermes D, Balakrishnan AV, editors. Stochastic differential systems. Springer Berlin Heidelberg; 1981, p. 25–40.
[49] Birge JR. The value of the stochastic solution in stochastic linear programs with fixed recourse. Math Program 1982;24(1):314–25.
[50] Bali TG. An extreme value approach to estimating interest-rate volatility: Pricing implications for interest-rate options. Manag Sci 2007;53(2):323–39.
[51] Bertsimas D, Doan XV, Natarajan K, Teo C-P. Models for minimax stochastic linear optimization problems with risk aversion. Math Oper Res 2010;35(3):580–602.
[52] Bertsimas D, Shtern S, Sturt B. A data-driven approach to multistage stochastic linear optimization. Manag Sci 2023;69(1):51–74.
[53] Bierlaire M, Fadda E, Tiotsop LF, Manerba D. A chance-constraint approach for optimizing social engagement-based services. 2022, p. 301–4.
[54] Bierlaire M, Fadda E, Tiotsop LF, Manerba D. Workforce allocation for social engagement services via stochastic optimization. Stud Comput Intell 2024;1158 SCI:71–87.