Stochastic Optimization in Charging Station Planning and a Robust Methodology with a Statistical Upper Bound Framework

by

Shixin Liu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Systems Engineering)
in the University of Michigan-Dearborn
2024

Doctoral Thesis Committee:

Associate Professor Jian Hu, Chair
Associate Professor Armagan Bayram
Professor Armen Zakarian
Principal Scientist Zhi Zhou, Argonne National Laboratory

Shixin Liu

liushixi@umich.edu

ORCID iD: 0000-0002-9579-7707

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Metropolitan areas globally are increasingly adopting electric taxis (ET) to mitigate transportation-related emissions, leading to a significant impact on urban transportation networks (TN) and electricity power distribution networks (PDN). This evolution presents challenges, including higher electricity demand and altered demand profiles, which underscore the critical interdependence between these systems. To address these complexities, we propose a novel two-stage stochastic programming planning model aimed at optimizing both the TN and PDN. This model strives to balance ET drivers' charging preferences, minimize infrastructure deployment and grid expansion costs, and ensure synchronized coordination between the TN and PDN. In transitioning to a more sustainable urban mobility paradigm, we also consider the advantages of incorporating an autonomous ET fleet to boost system efficiency.

The application of stochastic optimization in this context, however, is often obstructed by distributional ambiguity - where crucial probability distributions are not well-defined or are unknown. To counteract this, our study introduces an innovative approach that leverages the Average Percentile Upper Bound (APUB) framework. APUB is designed to minimize a statistical upper bound for the expected value of uncertain objectives, thus providing a more robust foundation for decision-making under uncertainty. This method offers a statistically rigorous upper limit for the population mean while serving as a viable risk metric for the sample mean. By integrating APUB into our stochastic optimization model, we address distributional ambiguities inherent in the deployment of electric taxis and their charging stations, thereby enhancing the model's reliability, consistency, and comprehensibility. Our empirical investigations, using two-

stage product mix and multi-product newsvendor benchmark problems, demonstrate the effectiveness of the APUB-enhanced optimization framework compared to traditional methods such as sample average approximation and distributionally robust optimization. This combined approach ensures that the transition to electric taxis not only addresses environmental concerns but also proceeds with a robust and economically feasible strategy for infrastructure development.

Chapter 1 Introduction

As cities and policymakers focus more on sustainable transportation, the move towards electric taxi fleets is gaining momentum. This global shift to electric taxis (ETs) isn't just about modernizing transport-it's a crucial step towards cutting down greenhouse gas emissions and cleaning up urban air pollution. These efforts are key to achieving larger environmental goals and reducing the carbon footprint of the transport sector. As a result, cities around the world are increasingly incorporating electric vehicles into their public transit systems, aiming to create greener, more livable urban spaces. Despite the enthusiasm, the transition isn't without its hurdles. A major challenge is setting up the charging infrastructure needed to keep these taxis running. The effectiveness of electric taxis heavily relies on having an extensive, efficient, and seamlessly integrated network of charging stations ready to go.

The development of electric vehicle (EV) charging infrastructure involves more than just installing hardware; it requires intricate decision-making shaped by the interactions between the transportation network (TN) and the power distribution network (PDN). Planning and optimizing charging stations are crucial, as they affect the operational efficiency of ETs and their convenience for users. Additionally, the variable dynamics of charging demand, shaped by travel patterns, taxi operation modes, and urban layouts, complicate the planning process. These complexities call for sophisticated approaches that can manage the unpredictable nature of charging needs and the detailed interdependencies between TN and PDN.

In the broader realm of operations research and optimization, making decisions under uncertainty is a common challenge across various fields, including engineering and management.

Stochastic optimization offers a robust framework for dealing with this uncertainty by treating uncertain parameters as random variables. This method helps decision-makers prepare for multiple potential scenarios, enhancing the reliability and effectiveness of their decisions. However, applying stochastic optimization in real-world scenarios is often constrained by factors such as limited data, incomplete information, and complex system dynamics. These factors can lead to distributional ambiguity, where the exact probability distributions of random variables are unknown, further complicating decision-making.

This dissertation introduces an innovative framework that merges advanced stochastic optimization techniques with practical aspects of charging station planning in urban settings. By adopting a two-stage stochastic programming model that reflects the dynamics between TN and PDN, this study aims to provide a comprehensive strategy for efficiently deploying ET charging infrastructure. It accounts for the uncertainties in charging demands and operational conditions, proposing a systematic approach to infrastructure planning that promotes the sustainable expansion of electric mobility.

Moreover, this research explores methodological improvements in stochastic optimization to tackle distributional ambiguity, a prevalent issue in practical applications. Integrating statistical inference techniques into the optimization process enhances the framework's capability to handle uncertainty, thus improving the quality and outcomes of decision-making. This integration marks a significant advancement in optimization, offering a more robust and reliable tool for addressing the complex challenges of developing electric vehicle charging infrastructure.

As electric taxis become a practical alternative to traditional fuel-based taxis, effective planning for charging stations is increasingly important. The adoption of ETs relies not just on the vehicles but also on the infrastructure that supports their operation. This dissertation contributes

to efforts aimed at enhancing urban transportation sustainability by offering a flexible and comprehensive framework for planning charging stations. Through detailed analysis and optimization, this work addresses the challenges posed by taxi fleet electrification, aiding a smoother transition to electric mobility and supporting broader environmental and urban quality goals.

Chapter 2 A Stochastic Charging Station Deployment Model for Electrified Taxi Fleets in Coupled Urban Transportation and Power Distribution Networks

Nomenclature

Sets

$\mathcal{I}$	Set of the nodes in the TN which represent geographical zones.
$\mathcal{N}$	Set of the nodes in the PDN.
$\mathcal{E}$	Set of the edges in the PDN.
$\mathcal{E}^+$	$\mathcal{E} \cup \{(0,1)\}$ where edge $(0,1)$ connects the dummy node 0 and the substation at node 1 and is used when increasing the capacity of the substation.
$\mathcal{J}$	Set of the candidate locations for constructing public charging stations.
$\mathcal{J}^+$	$\mathcal{J} \cup \{0\}$ where node 0 in the TN is a dummy charging station.
$\theta(\mathcal{J})$	Set of the nodes in the PDN which the candidates are connected with to get electrical power.
Ω	Sample space of the random scenario in the second stage of the model.
$\mathcal{B}$	Set of the constraints in the second stage under a scenario.

Parameters - TN

A_j Fixed cost of constructing a charging station at location j (\$/Station) ($A_j =$ \$163K for j in uptown and $A_j = \$652K$ for j in downtown).

C_j Unit construction cost of the charging capacity at location j where the capacity means the maximum number of the ETs accommodated to charge $(\$/ET)$ ($C_j = \$3{,}160$ for j in uptown and for $C_j = \$12{,}640\, j$ is in downtown).

$D_i(\omega)$ Charging demands in zone i under scenario ω (follow the patterns of charging demands in NYC).

α Construction cost coefficient.

Parameters - PDN

G_{mn} Cost of adding one distribution line on edge (m,n)(\$/ Line)$(G_{mn} = \$300K /$ Line for all (m,n))

L Unit cost of substation expansion $(\$/kVA)(L = \$788/kVA)$.

W Real power needed to charge an ET$(kW)(W = 7.7kW)$.

K Maximum number of the distribution lines allowed to add on each edge ($K = 2$).

$P_n^{\text{load}}(\omega)$ Active power load at the PDN node n under scenario $\omega(kW)$ (follow IEEE33).

$Q_n^{\text{load}}(\omega)$ Reactive power load at the PDN node n under scenario $\omega(kW)$ (follow IEEE33).

P_{mn}^{max} Capacity of the PDN edge (m,n) for the active power flow (kW) (follow IEEE33).

Q_{mn}^{max} Capacity of the PDN edge (m,n) for the reactive power flow (kvar) (follow IEEE33).

$V^{\max}/V^{\min}$ Upper/Lower bound of the voltage magnitude required for preserving good electric power quality $(p.u.)$ $(V^{\max}/V^{\min}=0.95/1.05\,p.u.)$.

R_{mn} Resistance of a distribution line on the PDN edge $(m,n)(\Omega)$ (follow IEEE33).

X_{mn} Reactance of a distribution line on the PDN edge $(m,n)(\Omega)$ (follow IEEE33).

β Expansion cost coefficient.

$\theta(j)$ The node in the PDN with which station j is connected to get electrical power.

$\theta^{-1}(n)$ The station which is connected with node n in the PDN to get electrical power.

Parameters - Penalty and Preference

T Unit monetary value of a driver's satisfaction with charging ($\$K$/ unit satisfaction) $(T=\$30K/\text{ unit satisfaction})$.

H_0 Penalty cost incurred when a driver cannot find any charging stations acceptable according to his or her preference $(\$K/ET)(H_0=1.5T)$.

H_1 Penalty cost incurred when a driver's charging request is rejected because the charging capacity of his or her selected charging station has overflowed $(\$K/ET)(H_1=2T)$.

$U_{ij}(\omega)$ A driver's preference to drive from zone i to station j for charging the ET under scenario ω.

$b(\omega)$ Coefficient indicating the sensitivity of an ET driver's preference to the travel time to a charging station.

$c(\omega)$ Coefficient representing the sensitivity of an ET driver's preference to the charging price.

$U^{\min}(\omega)$ Threshold of the charging preference under scenario ω below which a driver quits his or her charging need ($U_{\min}(\omega) = 0.35$ during the daytime and $U_{\min}(\omega) = 0.25$ at night).

ΔU Level of centralized charging management and vehicular automation.

Decision Variables - First Stage

z Vector consisting of $z_j \in \{0,1\}$ indicating if location $j \in \mathcal{J}$ is selected to construct a charging station.

x Vector consisting of x_j indicating charging capacity of station $j \in \mathcal{J}$.

u Vector consisting of u_{01} indicating the amount of substation expansion and $u_{mn} \in \{0,\ldots,K\}$ indicating the number of the distribution lines added onto the PDN edge $(m,n) \in \mathcal{E}$.

Decision Variables - Second Stage

s_j Charging requests rejected at station j.

y_{ij} Charging demands in zone i assigned to station j.

p_{mn} Active power flow through the PDN edge (m,n).

q_{mn} Reactive power flow through the PDN edge (m,n).

V_m Voltage magnitude at the PDN node m.

2.1 Introduction

Electrification in the transportation industry has made rapid development Martinez et al. (2016). The fast-growing demand for decarbonizing travel motivates taxi and ride-hailing electrification in the last decade. All-electric taxi (ET) fleets have been deployed and started to provide services in Shenzhen and Taiyuan, two large cities in China, as a consequence of China's policies for promoting taxi electrification Li et al. (2020), Guan et al. (2020). Ohio, New York City, and Charlotte in the U.S. successively added electric vehicles to their taxi fleet NYC TAXI LIMOUSINE COMMISSION (2013), Hirschfeld (2019), Plautz (2018). The United State Administration's goal is also to accelerate the electrification of transportation The White House (2021) and the deployment of charging infrastructure House (2021). ETs are capable of significantly cutting down operational costs and highly potentially improving urban air quality when deployed intensively in populated areas; however, their driving range limitations and high reliance on power system infrastructure are found to be critical barriers to large-scale adoption Egbue and Long (2012), Hagman and Langbroek (2019) Moreover, considering that thousands of cabdrivers cruise metropolitan road networks looking and waiting for passengers under uncertainty and they rely on charging on road heavily Cai et al. (2014), taxi electrification should involve the grid upgrade and co-optimization of the transportation network (TN) and the power distribution network (PDN) Wei et al. (2019). Indeed, a large number of ETs could add an overload of stress onto an existing PDN. Once the charging demands of ETs overreach electricity delivery capacity, not only a lot of demands are missed but also long waiting lines for charging could aggravate traffic congestion. Unfortunately, the major power grids in the United States, deployed in the early 1930s, are under pressure to meet the rising charging demands El-Hawary (2008), Nelles (1984).

Taxis empower a public transport chain with 24-hour/365-day availability, coupled with door-to-door customized service. Statistical evidence shows that in some cases (e.g., night, airport, people with disabilities, and sparsely populated rural areas), taxis are the unique or preferred public mobility service provider. Such flexibility makes the charging demands of the ET fleet vary spatially and temporally. Particularly, the charging demands are highly uncertain, on account of the random distribution of taxi supply and passenger needs. Obviously, an effective and efficient strategy for deploying public charging stations is essential to maintain the productivity and feasibility of taxi services after electrification.

Over the last decades, the reliance on the TN and PDN has been discussed in the problem of electric vehicle charging station placement. The studies in Arias et al. (2017), de Quevedo et al. (2017), Wang et al. (2019) make joint decisions on allocating charging stations and upgrading PDN assets such as distribution lines, substations, and other distributed energy resources. Wang et al. (2018b) further takes into consideration the expansion of the TN described with traffic flows in an unconstrained traffic assignment mode. Recently, Deb et al. (2021), Ehsan and Yang (2019) propose stochastic programming models to consider operations in either the TN or the PDN under uncertainty. Such those studies are targeted at residential or commercial electrical vehicles with regular routes. Accordingly, transportation networks are modeled using origin-destination patterns. However, given ETs' distinct operational characteristics and dynamic charging needs, this method is not fit for predicting their routing and charging behavior.

The involvement of ETs in charging station planning has caught the attention recently addressed TN, PDN, or ET charging behavior separately. Kaya et al. (2020) utilize GIS information, fuzzy analytic hierarchy process, and order preference by similarity to an ideal solution. Meng et al. (2020) optimize the social costs including travel costs, investment costs, and

waiting costs. Cilio and Babacan (2021) model a datadriven framework using an iterative clustering technique to determine the number of charging stations and a numerical optimization method to find their best locations. Zhang et al. (2019) cluster the k-means distributions of ETs and use the barycentric approach to select locations of charging stations. Furthermore, Clairand et al. (2022) describe deterministic requirements on both the TN and PDN. In addition, contrary to the emphasis on charging station siting, Sossan et al. (2020) compare the effects of human and self-driven electric vehicles charging on the PDN. They observe that autonomous vehicles can select the most suitable charging locations to reduce power congestion and consequently save nearly half of the charging time needed by humandriven vehicles. Bauer et al. (2021) exhibit that real-time information sharing is an alternative effective way to promote charging efficiency.

The existing literature on charging station planning for ETs has predominantly focused on the aspects of TN, PDN, or ET charging behavior in isolation. In contrast, our study, conducted on behalf of city managers, explores the intricate interplay between TN and PDN in relation to ET drivers' charging behavior. We introduce a two-stage stochastic programming framework that integrates the long-term planning phase, involving the deployment of charging stations within the TN and the expansion of power distribution lines in the PDN, with the operational phase that accounts for the inherent uncertainty in ET charging services. In project planning, it is essential to account for the entire time horizon. The operational phase of our model can explore various scenarios occurring over the entire time span of the project. These scenarios represent different traffic patterns and power consumption conditions during specific time periods (including peak and non-peak hours, daytime and nighttime, weekdays and weekends), as well as seasonal variations (e.g., summer and winter). Our recommended solution is designed to optimize charging service performance across a wide range of operational conditions. To the best of our knowledge,

no existing literature addresses the interconnected TN and PDN while considering the stochastic nature of ET charging, especially in the context of varying operational conditions during different time periods.

We also analyze the different effects of deploying human-driven and autonomous ET fleets in the charging station planning problem. Human drivers select charging stations according to their individual preferences. It is highly possible, in populated areas, that there are much more competitors than expected swarming into the same charging station on account of lower charging prices, shorter distances, etc. As a consequence, demands on the local electrical grid may spike without proper management. In comparison with human drivers, autonomous ETs are less selfish. Hence, fleet operators can easily maneuver and coordinate charging services for autonomous ETs to optimize the allocation of electric power. In the literature electric vehicle users' charging preferences are considered only in a few works: Wang et al. (2018a) considers reducing the excessive waiting time but still, there is a negative impact on the users' experience. User convenience is maximized as an objective in the charging selection process in research Wen et al. (2012), Malhotra et al. (2016), Xia et al. (2016), Chung et al. (2018). User convenience is quantified in Malhotra et al. (2016) with respect to remaining SOC, remaining time to charge, and charging rate. In Chung et al. (2018), a parameter is defined for user convenience where EVs need less electricity and are close to plug-off time should have higher charging priority.

Our contributions are summarized as follows:

Comprehensive coverage of scenarios: our modeling framework distinguishes between a long-term project planning phase and an operational phase. One notable advantage of our approach lies in the comprehensive coverage of scenarios throughout the project's entire duration. During

the operational phase, we explore a range of scenarios, each representing different traffic patterns and power consumption conditions occurring at specific times.

Innovative use of a mixed probability distribution: what sets our work apart in the literature is our innovative use of a mixed probability distribution to capture the stochastic nature of these scenarios. Within this framework, the mixture weight denotes the relative importance assigned to each class of time period, as determined by subjective probabilities established in collaboration with project stakeholders. In each scenario, we leverage the steady-state probability distribution of random traffic and electric consumption conditions.

Scenario-specific preference representation: we are the first to introduce a scenario specific preference representation, grounded in the theory of state-dependent utility, to elucidate how drivers' preferences and satisfaction are contingent upon circumstances varying across different time periods.

Analysis human-driven and autonomous ET fleets: we model and analyze the different effects of deploying human-driven and autonomous ET fleets in the charging station planning problem.

This chapter is organized as follows. In Section 2.2 we present the two-stage stochastic programming model. The first stage described in Section 2.2.1 considers a long-term planning strategy of constructing charging stations and expanding the PDN and the second stage in Section 2.2.2 seeks an optimal policy to serve ET charging demands during a steady-state operating period. Section 2.2.3 reformulates the mixed-integer nonlinear model to a mixed-integer linear program. We conduct a numerical study in Section 2.3. Section 2.4 concludes this paper.

2.2 Mathematical Model

We now discuss the two-stage stochastic programming model in Figure 2.1. In a longterm planning stage, we upgrade the PDN to increase its electrical load capacity and, on this basis, to determine the locations and sizes of charging stations for boosting electricity supply for charging demands. In a steady-state operating stage, we characterize ET drivers' preference-specific charging behavior, model the electricity delivery, and depict the inherent randomness of ET charging ascribed to the variations of charging demand and traffic condition in the TN and electricity consumption and price in the PDN. A node (called T-node) in the TN represents a geographical zone, while a node (called P-node) in the PDN is a "junction box" where two or more distribution lines meet. The locations of charging stations are selected among the candidate T-nodes (denoted by $\mathcal{J}$). All these candidates, if selected, can be connected to the PDN at corresponding P-nodes. Assume that there is a one-to-one correspondence, denoted by θ, between the candidates and their connected P-nodes. Then, as described in the nomenclature, $\theta(j)$ is the P-node which T-node j is connected with, $\theta(\mathcal{J})$ is the set of all the P-nodes connected with candidates, and $\theta^{-1}(n)$ is the T-node which is connected with P-node n. This joint relationship is used to model the second stage.

2.2.1 The First Stage of the Model

The first stage of the model is to balance the costs of expanding the PDN and constructing charging stations and ET drivers' dissatisfaction with charging. The cost of constructing charging stations is made up of the fixed setup cost $\left(\sum_{j\in\mathcal{J}} A_j z_j\right)$ and the cost related to their charging capacities $\left(\sum_{j\in\mathcal{J}} C_j x_j\right)$ as follows:

$$\tau(\boldsymbol{z},\boldsymbol{x}) = \alpha\left(\sum_{j\in\mathcal{J}} A_j z_j + \sum_{j\in\mathcal{J}} C_j x_j\right), \tag{2.1}$$

where α, called the construction cost coefficient, embodies economic considerations. A larger α means a higher unit cost to construct a charging station. The reason of using α in the model is that the case study in Section 2.3 evaluates the impact of the

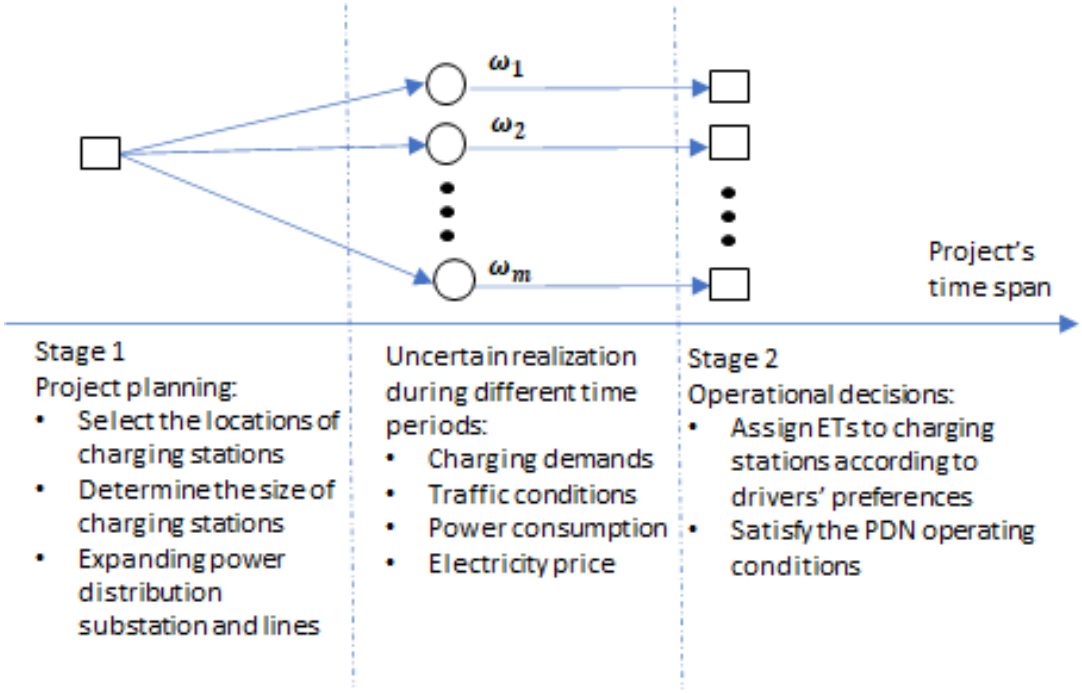


Figure 2.1: Two-Stage Stochastic Programming Model

expansion cost by varying α. To upgrade the PDN, we consider the costs of expanding its substation (Lu_{01}) and adding new distribution lines $\left(\sum_{(m,n)\in\mathcal{E}} G_{mn}u_{mn}\right)$ adjusted by β as:

$$\phi(\boldsymbol{u}) = \beta\left(Lu_{01} + \sum_{(m,n)\in\mathcal{E}} G_{mn}u_{mn}\right), \tag{2.2}$$

where β is the expansion cost coefficient indicating the unit cost of expanding the PDN. Similarly, β is used in the case study to test the impact of the expansion cost. The substation expansion

increases the flow capacity at the root node of the PDN, which receives power from a transmission network while adding extra distribution lines allows higher flow to pass corresponding edges in the PDN.

The first-stage model is expressed as follows:

$$\min_{\boldsymbol{z},\boldsymbol{x},\boldsymbol{u}} \tau(\boldsymbol{z},\boldsymbol{x}) + \phi(\boldsymbol{u}) + \mathbb{E}\left[\Pi(\boldsymbol{z},\boldsymbol{x},\boldsymbol{u},\omega)\right] \tag{2.3a}$$

$$\text{s.t.} B_j z_j \le x_j \le \tilde{M}_j z_j, \qquad j \in \mathcal{J}, \tag{2.3b}$$

$$z_j \in \{0,1\}, \qquad j \in \mathcal{J}, \tag{2.3c}$$

$$x_j \ge 0, \qquad j \in \mathcal{J}, \tag{2.3d}$$

$$u_{01} \ge 0, \tag{2.3e}$$

$$u_{mn} \in \{0,1,\cdots,K\}. \tag{2.3f}$$

In the objective (2.3a), $\mathbb{E}\left[\Pi(\mathbf{z},\mathbf{x},\mathbf{u},\omega)\right]$ measures the steady-state expectation (or called long-run average) of ET drivers' dissatisfaction with charging measured by a monetary value. The random scenario $\omega \in \Omega$ (Ω is a sample space) accounts for the uncertainties in charging demand, traffic time, electricity consumption, and electricity price in the operational stage introduced in 2.2.2. In project planning, it is essential to account for the entire time horizon. In our model, Ω encompasses scenarios that cover specific classes of time period, including peak and non-peak hours, daytime and nighttime, weekdays and weekends, as well as seasonal variations, such as summer and winter. It's crucial to recognize that traffic patterns and power consumption conditions can vary significantly during these different classes of time period. The second stage of our model conducts a comprehensive evaluation of a potential solution of the first stage, taking into consideration a wide range of operational conditions across various scenarios. Accordingly, we employ a mixed probability distribution to account for the random nature of the scenario ω, i.e.

$\omega \sim \sum \alpha_t F_t$, where the weight α_t indicates the relative importance of the t-th class of time period, determined by project stakeholders, and F_t is the steady-state probability distribution of random traffic and electric consumption conditions in t. Notably, based on the steady-state distribution, $\mathbb{E}\left[\Pi\left(\mathbf{z}, \mathbf{x}, \mathbf{u}, \omega\right)\right]$ represents the long-run average of ET drivers' dissatisfaction.

Constraint (2.3b) indicates that the charging capacity $x_j = 0$ if the candidate j is not selected and otherwise $B_j \le x_j \le \tilde{M}_j$. Here, B_j gives the smallest size of a charging station required to build at location j and $\tilde{M}_j$ is the largest possible charging demand over all scenarios at the second stage as

$$\tilde{M}_j = \sup_{\omega \in \Omega} \left\{ M_j\left(\omega\right) \right\}.$$

The detailed explanation of $M_j\left(\omega\right)$ is addressed when describing the second stage. Constraints (2.3c)-(2.3f) depict the denotations of the decision variables $\mathbf{z}, \mathbf{x}$, and $\mathbf{u}$ given in the nomenclature.

2.2.2 The Second Stage of the Model

The second-stage objective is to minimize the penalty for unmet charging needs and maximize ET drivers' satisfaction with charging under some conditions describing drivers' preference-specific charging behavior and distribution grid operation. There are two penalty costs denoted by $h_0\left(\mathbf{s}\right)$ and $h_1\left(\mathbf{s}\right)$. The former is incurred when some drivers are not willing to charge at any constructed stations because of long travel times to the stations or high charging prices at that moment, while the latter is ascribed to the demands beyond the charging capacities of the

stations. Denote by $g(\mathbf{y},\omega)$ the measure function of ET drivers' satisfaction under the random scenario ω.

The second-stage model is written as

$$\Pi(\mathbf{z},\mathbf{x},\mathbf{u},\omega)=\min_{(\mathbf{s},\mathbf{y})\in\mathcal{B}(\mathbf{z},\mathbf{x},\mathbf{u},\omega)} h_0(\mathbf{s})+h_1(\mathbf{s})-g(\mathbf{y},\omega), \tag{2.4}$$

where the set $\mathcal{B}(\mathbf{z},\mathbf{x},\mathbf{u},\omega)$ consists of all the conditions on drivers' charging behavior and the distribution grid parameterized by the decisions made in the first stage. Without loss of generality, we limit our discussion to a certain scenario in the later statement.

Assume that human drivers merely select charging stations that they prefer over the others and the charging preference is related to the charging price at a charging station and the travel time to there Yang [2018]. We characterize the preference as an exponential preference function of the charging price and travel time as follows:

$$U_{ij}(\omega)=\exp\left(-b(\omega)\eta_j(\omega)-c(\omega)\psi_{ij}(\omega)\right), \tag{2.5}$$

where $U_{ij}(\omega)$ measures the preference value of an ET driver in zone i selecting the station at location j. In economics and decision theory, exponential functions are commonly used to represent the preferences of individuals who exhibit constant absolute risk aversion (CARA). CARA means that the decision-maker's aversion to risk, which arises from uncertainty, is consistent across different preference levels. The charging price $\eta_j(\omega)$ and travel time $\psi_{ij}(\omega)$ both varies in scenarios, as do the coefficients $b(\omega)$ and $c(\omega)$. This scenario-specific preference representation, grounded in the theory of state-dependent utility Drèze and Rustichini [2004], elucidates how drivers' preference and satisfaction are contingent upon circumstances varying across different time periods. For instance, during the winter season, drivers may exhibit lower tolerance for extended travel times due to the reduced range of their electric vehicle battery.

Assume that drivers are reluctant to select the stations where the preference values of charging are lower than a given threshold $U^{\min}(\omega)$. That is, drivers in zone i do not select station at location j if $U_{ij}(\omega) < U^{\min}(\omega)$. We say that zone i is covered if there exists at least $j \in \mathcal{J}$ such that $U_{ij}(\omega) \geq U^{\min}(\omega)$ and otherwise uncovered. Suppose that the demands in uncovered zones are served at a dummy station with $j = 0$ with the incurred penalty

$$h_0(\mathbf{s}) = H_0 s_0, \tag{2.6}$$

where s_0 counts all the demands served at the dummy station.

Oppositely, drivers in covered zones select certain stations to charge. The preference values of these selections are converted to monetary value by a coefficient T and counted into drivers' satisfaction with charging as

$$g(\mathbf{y}, \omega) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} T U_{ij}(\omega) y_{ij}, \tag{2.7}$$

However, these drivers face two possible situations after selecting stations to charge. Some demands can be served, others may be rejected since the total demands toward some stations are beyond their charging capacities. For the rejected demands, we assign the second penalty as

$$h_1(\mathbf{s}) = \sum_{j \in \mathcal{J}} H_1 s_j. \tag{2.8}$$

We next describe the set $\mathcal{B}(\mathbf{z}, \mathbf{x}, \mathbf{u}, \omega)$ of the constraints.

In the domain of electric vehicle charging models, a common practice involves the utilization of origin-destination (OD) patterns. However, it is important to note that these patterns are not directly applicable to ET fleets. Existing literature offers a variety of spatiotemporal analysis methods for forecasting taxi demand Yao et al. [2018], Rodrigues et al. [2019], Askari et al. [2020], Gangrade et al. [2022], typically entailing the division of a city into distinct zones to

predict time-dependent demand within each zone. In alignment with this convention, our approach assumes that ET charging demands are randomly generated at the nodes of a transportation network, which we consider as the central points of these zones in our paper. Our approach holds practical advantages as it allows us to leverage input data from existing spatiotemporal taxi forecast models for estimating ET charging demand. Moreover, it aligns well with our other assumption that ET charging demands exhibit temporal variation in specific scenarios.

To address the assignment of ET charging demands at network nodes under random scenarios, we describe drivers' charging behavior in the TN as follows:

$$0 \le \sum_{i \in \mathcal{I}} y_{ij} - s_j \le x_j, \qquad j \in \mathcal{J}^+, \tag{2.9}$$

$$0 \le s_j \le M_j(\omega) z_j, \qquad j \in \mathcal{J}^+, \tag{2.10}$$

$$\sum_{j \in \mathcal{J}^+} y_{ij} = D_i(\omega), \qquad i \in \mathcal{I}, \tag{2.11}$$

$$U_{ij}(\omega) y_{ij} \ge y_{ij} max\{U_{ik}(\omega) - \Delta U, U^{min}(\omega)\} - D_i(\omega)(1 - z_k), i \in \mathcal{I}, j \in \mathcal{J}^+, k \in \mathcal{J}^+, \tag{2.12}$$

$$y_{ij} \ge 0, \qquad i \in \mathcal{I}, j \in \mathcal{J}^+, \tag{2.13}$$

Constraints (2.9) and (2.10) state the relationship between the flow y and unmet demand s. If location j is not selected to construct a station in the first stage, s_j and y_{ij} should equal to 0 for all $i \in \mathcal{I}$; otherwise, constraint (2.9) requires that the demands served at each station should not exceed its charging capacity. Note that it is a dummy station at $j = 0$ and the demands in the zones uncovered by any stations flow to the dummy station. In the model we set $z_0 = 1, x_0 = 0$, and $U_{i0}(\omega) = U^{\min}(\omega), i \in \mathcal{I}$. In addition,

$$M_j(\omega) = \sum_{i \in \mathcal{I}_j(\omega)} D_i(\omega)$$

where $\mathcal{I}_j(\omega) = \left\{ i \in \mathcal{I} : U_{ij}(\omega) \geq U^{\min}(\omega) \right\}$ counts all the demands in the zones covered by the station at location j. In constraint (2.11) the demands in zone i are assigned to different stations. Analogous to the user-choice assignment model in Yang [2018], constraint (2.12) describes how drivers select stations to charge. The user-choice assignment model requires that a driver should select a unique station to charge where the driver has the largest preference value. Differently, constraint (2.12) sets $\Delta U \in [0,1]$ to quantify the extent where drivers allow for sacrificing their preferences, and our model can optimally assign charging demands to boost the systematic performance, i.e., the second-stage objective function in (2.4). We thus call ΔU the level of centralized charging management and vehicular automation. For example, ΔU should be zero or a very small number for myopic human drivers since they are not pleased to compromise their individual best choices for improving the systematic performance. When $\Delta U = 0$, constraint (2.12) is the same as the user-choice assignment model. ETs at zone i select the charging station j, i.e., y_{ij} is strictly positive, if and only if the ET's preference value $U_{ij}(\omega)$ at station j is larger than any other station $U_{ik}(\omega)$, including the dummy station $U^{\min}(\omega)$. In contrast, autonomous ETs are less selfish, and correspondingly ΔU is 1, for which the ETs may not select the station which they are most satisfied with. Subsequently, the distribution system operators in the PDN can easily maneuver and coordinate autonomous ETs' charging requests.

In what follows we consider the distribution system operation constraints in the PDN.

$$\sum_{m:(m,n)\in\mathcal{E}^+} p_{mn} - \sum_{k:(n,k)\in\mathcal{E}^+} p_{nk} = P_n^{load}(\omega) + W\left(\sum_{i\in\mathcal{I}} y_{i,\theta^{-1}(n)} - s_{\theta^{-1}(n)} \right), \quad n \in \theta(\mathcal{J}), \tag{2.14}$$

$$\sum_{m:(m,n)\in\mathcal{E}^+} p_{mn} - \sum_{k:(n,k)\in\mathcal{E}^+} p_{nk} = P_n^{\text{load}}(\omega), n \in \mathcal{N} \setminus \theta(\mathcal{J}), \tag{2.15}$$

$$\sum_{m:(m,n)\in\mathcal{E}^+} q_{mn} - \sum_{k:(n,k)\in\mathcal{E}^+} q_{nk} = Q_n^{\text{load}}(\omega), n \in \mathcal{N}, \tag{2.16}$$

$$V_m^2 - V_n^2 = \frac{2R_{mn}p_{mn} + 2X_{mn}q_{mn}}{1+u_{mn}}, (m,\ n) \in \mathcal{E}, \tag{2.17}$$

$$V_0 = 1 p.u., \tag{2.18}$$

$$V^{\min} \le V_n \le V^{\max},\ \ n \in \mathcal{N}, \tag{2.19}$$

$$0 \le p_{mn} \le P_{mn}^{\max}(1+u_{mn}), (m,\ n) \in \mathcal{E}, \tag{2.20}$$

$$0 \le p_{01} \le P_{01}^{\max} + u_{01}, \tag{2.21}$$

$$0 \le q_{mn} \le Q_{mn}^{\max}(1+u_{mn}), (m,\ n) \in \mathcal{E}, \tag{2.22}$$

$$0 \le q_{01} \le Q_{01}^{\max} + u_{01}. \tag{2.23}$$

Constraints (2.14)-(2.16) formulate the active and reactive power flows to satisfy existing electricity load and ET charging demands. Only active charging power from ETs is considered with the assumption that most ETs are not equipped with smart inverters and do not participate in providing reactive power. Constraint (2.17) is the voltage drop equation with u_{mn} extra lines on the edge (m,n) Guo et al. [2019]. Constraint (2.18) sets the voltage at a transmission substation as 1 p.u., which means the voltage at the substation is fixed at the desired voltage level. Constraint (2.19) sets the upper and lower limit of node voltage, which is usually 1.05 p.u. and 0.95 p.u. according to ANSI standards to ensure power quality is delivered to customers. Constraints (2.20) and (2.22) are active and reactive power flow limits for all lines considering line expansion. Constraints (2.21) and (2.23) are power flow boundaries for the substation considering expansion.

This set $\mathcal{B}(\mathbf{z},\mathbf{x},\mathbf{u},\omega)$, integrating the conditions described above, consists of all points (s,y) satisfying conditions (2.9) - (2.23) for given $\mathbf{z},\mathbf{x},\mathbf{u}$, and ω.

2.2.3 Tractable Reformulation and Solution Method

The aforementioned model aggregates the constraints related to the TN and PDN. Either of the two networks is a large-scale complex system. Moreover, the uncertainties on charging demand, traffic time, electricity consumption, and electricity price considered in the model greatly increase the size of the model. In this section, we discuss a solution approach to linearize the model and reduce redundant constraints.

Linearization. The second-stage model (2.4) has the nonlinear constraint (2.17). To linearize this constraint, let the number of expansion lines $u_{mn}=\sum_{k=1}^{K}k\tilde{u}_{mn}^{k}$ where $\tilde{u}_{mn}^{k}\in\{0,1\}$. If there are exact k expansion lines added on edge $(mn),\tilde{u}_{mn}^{k}=1$; otherwise, $\tilde{u}_{mn}^{k}=0$. It implies that $\sum_{k=1}^{K}\tilde{u}_{mn}^{k}\le 1$ for all $(m,n)\in\mathcal{E}$. Also, referring to Yang et al. [2017] Yang et al. [2018], we let $v_m=V_m^2$ and replace constraints (2.18) and (2.19) with $v_0=1p.u.$ and $\left(V^{\min}\right)^2\le v_n\le\left(V^{\max}\right)^2$ for $n\in\mathcal{N}$, respectively. Meanwhile, constraint (2.17) is reformulated as

$$(v_m-v_n)+\sum_{k=1}^{K}k\tilde{u}_{mn}^{k}(v_m-v_n)=2R_{mn}p_{mn}+2X_{mn}q_{mn},\quad (m,n)\in\mathcal{E}.$$

Introducing a new decision variable $r_{mn}^{k}=\tilde{u}_{mn}^{k}(v_m-v_n)$, we obtain

$$
\begin{aligned}
&(v_m - v_n) + \sum_{k=1}^{K} k r_{mn}^k = 2R_{mn} p_{mn} + 2X_{mn} q_{mn}, && (m,n) \in \mathcal{E}, \\
&0 \le r_{mn}^k \le \left(V^{\max}\right)^2 \tilde{u}_{mn}^k, && (m,n) \in \mathcal{E}, \\
& && k \in \{1, \cdots, K\}, \\
&0 \le (v_m - v_n) - r_{mn}^k \le \left(V^{\max}\right)^2 \left(1 - \tilde{u}_{mn}^k\right) && (m,n) \in \mathcal{E}, \\
& && k \in \{1, \cdots, K\}.
\end{aligned}
$$

Constraint reduction. The size of the second-stage model (2.4) is mainly determined by the number of the constraints defined in (2.12), which is $|\mathcal{I}| \times |\mathcal{J}^+|^2$. We next discuss how to reduce the number of constraints. Constraint (2.12) is redundant if $U_{ij}(\omega) \ge \max\{U_{ik}(\omega) - \Delta U, U^{\min}(\omega)\}$; otherwise, it is equivalent to the inequality $y_{ij} \le D_i(\omega)(1 - z_k)$. Hence, we replace constraint (2.12) with

$$
\begin{aligned}
&1\{U_{ij}(\omega) < \max\{U_{ik}(\omega) - \Delta U, U^{\min}(\omega)\}\}\, y_{ij} \\
&\quad \le D_i(\omega)(1 - z_k), i \in \mathcal{I}, j \in \mathcal{J}^+, k \in \mathcal{J}^+,
\end{aligned} \tag{2.24}
$$

where $1\{\cdot\}$ is the indicator function. Constraint (2.24) for $k \in \mathcal{J}^+$ is redundant if station k is not constructed $(z_k = 0)$ or $U_{ij}(\omega) \ge \max\{U_{ik}(\omega) - \Delta U, U^{\min}(\omega)\}$; otherwise, drivers in zone i prefer station k to station j with $U_{ij}(\omega) < \max\{U_{ik}(\omega) - \Delta U, U^{\min}(\omega)\}$ and thus $y_{ij} = 0$. By this observation, we can further reformulate (2.24) as

$$
\begin{aligned}
&\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^+} \mathbb{1}\{U_{ij}(\omega) < \max\{U_{ik}(\omega) - \Delta U, U^{\min}(\omega)\}\}\, y_{ij} \\
&\quad \le (1 - z_k) \sum_{i \in \mathcal{I}} D_i(\omega), k \in \mathcal{J}^+.
\end{aligned} \tag{2.25}
$$

Reformulated Model.

Solution method. Considering that the first-stage (2.3) in the model is MILP and the second-stage is LP, the classical integer L-shaped method Laporte and Louveaux [1993] can be well applied. The iOptimize solver Huang [2016], embedded with the integer L-shaped method, is used in the numerical study of Section 2.3.

If the sample space Ω is uncountable or exceedingly large, we employ the Sample Average Approximation (SAA) method - a widely-used technique for addressing stochastic programming problems (see Shapiro et al. [2009] and references therein) to solve the model. SAA approximates the expected value $\mathbb{E}\left[\Pi\left(\mathbf{z},\mathbf{x},\mathbf{u},\omega\right)\right]$ using the sample mean of a smaller number of independent and identically distributed (i.i.d.) observations generated from the sample space Ω. Its asymptotic convergence ensures solution quality for the approximation. For a comprehensive understanding of SAA's asymptotic convergence and its convergence rate, we recommend referring to Shapiro et al. [2009] where these topics are elaborated in detail.

In addition to SAA, we employ a warm start process to manage the computational complexity associated with a large number of random scenarios. This strategic approach involves solving the problem initially using a limited number of i.i.d. resampled scenarios. The key benefit lies in utilizing the optimal solution derived from this simplified scenario as the initial values for the full problem. This process proves particularly effective when dealing with stochastic MILP problems. By adhering to the SAA principle, the warm start process ensures the quality of the initial solution, even when based on a limited number of scenarios. Consequently, this approach significantly reduces computational costs. In Section 2.3 we numerically illustrate the effectiveness of the warm start process.

2.3 Numerical Analyses

This case study evaluates the performance of the charging station planning model, tests the impact of the costs of expanding the PDN and constructing charging stations, and compares the effects of human-driven and autonomous ET fleet. We use the Sioux-Falls transportation network Baran and Wu [1989] in Figure 2.2a and the IEEE 33-bus test feeder Leblanc [1975] in Figure 2.2b as the TN and PDN in the analyses of uncertainty and economic performance. Also, the Eastern Massachusetts (EMA) network Zhang et al. [2016] and IEEE 123-bus feeder DSA Subcommitte are involved in the computational analysis as the benchmark with a larger size.

There are eleventh candidate sites among which our model optimally selects the locations of charging stations. All the candidates are marked with the large dots in Figure 2.2a and Table 2.1 links them with the P-nodes in the PDN. The electrical supply for the charging station constructed at a selected candidate site needs to be provided by the PDN at the corresponding P-node. Figure 2.2b marks these P-nodes in the PDN.

Table 2.1: The mapping of the candidates in the TN and the connected P-nodes in the PDN

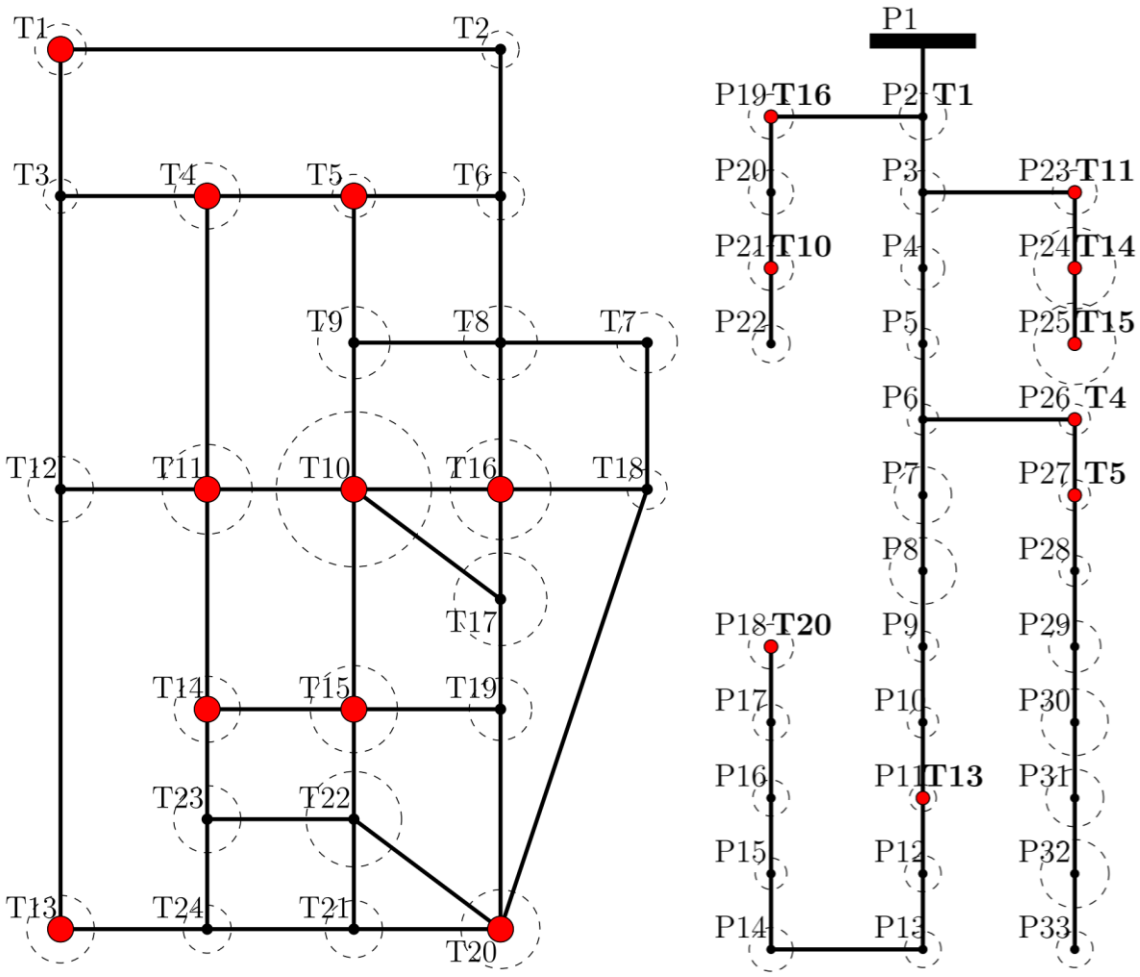
TN	T1	T4	T5	T10	T11	T13	T14	T15	T16	T20
PDN	P2	P26	P27	P21	P23	P11	P24	P25	P19	P18

Table 2.2: Partition of the TN and PDN

Districts	Abbreviation	TN Nodes	Distribution Lines
Eastern District	ED (CBD)	T7, T8, T9, (T10),(T16), T17, T18	P1-P22
Western District	WD	(T11), T12, (T14), (T15)	P1-P15
Northern District	ND	(T1), T2, T3, (T4), T5, T6	P1-P33
Southern District	SD	(T13), T19, (T20), T21, T22, T23, T24	P1-P18

The T-nodes in the parentheses are the candidate locations.

We partition the TN into four districts as shown in Table 2.2: eastern district (ED),



(a) Sioux Falls transportation network (b) IEEE 33-node

Figure 2.2: Test Networks

Table 2.3: Performance Indices

Index	Abbreviation	Description	Mathematical Expression
Expected number of charged ETs	Charged#	The expected number of charged ETs over all scenarios	$\mathbb{E}\left[\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} y_{ij}^{*}(\omega) - s_{i}^{*}(\omega)\right]$
Expected number of uncharged ETs	Uncharged#	The expected number of uncharged ETs over all scenarios	$\mathbb{E}\left[\sum_{j\in\mathcal{J}} s_{j}^{*}(\omega)\right]$
Total charging capacity constructed	CAP#	The sum of the capacities of all the constructed charging stations	$\sum_{j\in\mathcal{J}} x_{j}$
Monetary value of Satisfaction $(\$K)$	SAT$	The monetary measure of drivers satisfaction of charging	$-\mathbb{E}\left[\Pi\left(\mathbf{z}^{*},\mathbf{x}^{*},\mathbf{u}^{*},\omega\right)\right]$
Total construction cost $(\$K)$	TN$	Fixed construction cost and capacity cost for charging stations	$\tau\left(\mathbf{z}^{*},\mathbf{x}^{*}\right)$
Total expansion cost $(\$K)$	PDN$	The total expansion cost of transmission substation and distribution lines	$\phi\left(\mathbf{u}^{*}\right)$
Optimal value $(\$K)$	OPT$	The optimal value of the model	TN$ + PDN$ - SAT$
Total return on investment	ROI	The ratio of the monetary value of satisfaction over the total cost	$\text{SAT\$} / (\text{TN\$} + \text{PDN\$})$

western district (WD), northern district (ND), and southern district (SD). The electrical substation at the root (P1) of the PDN is located in ND. In this study, ED is the central business district (CBD) with a large number of charging demands supported by the power distribution line P1-P22, and the other three districts are uptown areas supported by P1-P18, P1-P15, and P1-P33, respectively. Note that P1-P18 is the longest line carrying electrical power from the north to the south. We generate scenarios from a mixed probability distribution combining diverse operational conditions during peak and non-peak hours. Assume that, during either of peak and non-peak hours, the charging demands $\left(D_i(\omega)\right)$, active and reactive power usages $\left(P_n^{\text{load?}}(\omega), Q_n^{\text{load?}}(\omega)\right)$ are normally distributed random variables. At each time period, their means refer to the hourly patterns of the charging demands in NYC NYC TAXI LIMOUSINE COMMISSION [2013] and the power usages obtained by adjusting the IEEE 33-bus benchmark data according to the PJM's power load profiles in commercial and residential areas PJM [2022]. In Figure 2.2a and Figure 2.2b, the areas of the dotted circles visualizes the expected charging demands and the expected power usages. We convert the benchmark traffic flow data in the Sioux-Falls benchmark to arc travel times according to the speed-density model Greenshields et al. [1935] and further vary them during either peak or non-peak hours. The obtained deterministic times are assumed to be the means of normally distributed random travel times. This data generation approach allows us to take into consideration the traffic congestion during peak hours.

2.3.1 Performance Indices

In the test, we report the expected number of charged ETs, the expected number of uncharged ETs, the monetary value of satisfaction, the total construction cost, the total expansion cost, and the optimal value of the model, which are described in Table 2.3. Note that, in the table,

$\mathbf{z}^*, \mathbf{x}^*$, and $\mathbf{u}^*$ are the optimal solution of the model in the first stage. $y_{ij}^*(\omega)$ and $s_i^*(\omega)$ are the optimal solution of the model for the decision variables y_{ij} and s_i in the second stage at scenario ω.

To investigate the interdependence between the TN and PDN, we also calculate the monetary value of satisfaction in each district in Table 2.2. For each scenario ω, the satisfaction value at TN-node $i \in \mathcal{I}$ is given (referring to (2.6) - (2.8)) as

$$\mathrm{SAT}_i(\omega) = -H_0 y_{i0}^*(\omega) + \sum_{j\in\mathcal{J}} TU_{ij}(\omega) y_{ij}^*(\omega) - \sum_{j\in\mathcal{J}} H_1 \left(s_j^*(\omega) \frac{y_{ij}^*(\omega)}{\sum_{k\in\mathcal{I}} y_{kj}^*(\omega)} \right). \tag{2.26}$$

In the above equation (2.26), the first item is the penalty incurred if T-Node i is uncovered by any charging station, the second item is the preference value of the drivers' satisfaction with charging, and the last item is the penalty if some charging demands at T-Node i are rejected to serve. In the last item, we assume that all the charging requests received by the charging station at T-Node j have an equally likely chance to be rejected. Thus, the satisfaction value of a district is the expected value of the sum of $\mathrm{SAT}_i(\omega)$ for all i belonging to this district. On this basis, we define the decomposed ROI of a district for integrating the costs into the analysis. The decomposed ROI of a district is the ratio of the satisfaction value of this district over the total cost of construction and expansion (TN$ + PDN$). Note that the sum of the decomposed ROIs of the four districts is equal to the total ROI described in Table 2.3.

2.3.2 Computational Analysis

The computation analysis is conducted using the two groups of benchmark networks. The first group is the Sioux-Falls and IEEE 33-bus, with which the model has 74 binary variables and 11 continuous variables in the first stage and 437 continuous variables in the second stage. The second

group tests the EMA and IEEE 123-bus which have 48 TN nodes and 20 charging station candidates. This case provides 256 binary variables and 21 continuous variables in the first stage and 1621 continuous variables in the second stage. The test runs with the 11th Gen Intel(R) Core(TM) i7-11700 processor and 32GB RAM.

Table 2.4: Computational Analyses

Detail	Metric	2	6	18	54	108	432
Sioux-Falls, IEEE 33-bus							
Performance Indices	Time (sec.)	4	13	20	44	89	331
	SAT$	9455	8727	8105	5747	5437	5406
	TN$	5503	5653	5651	5477	5333	5583
	PDN$	4897	5610	5590	5236	5193	5336
	OPT$	945	2536	3136	4966	5089	5533
Stage Variables	Binary	74					
	Continuous	11 (Stage 1), 437 (Stage 2)					
EMA, IEEE 123-bus							
Performance Indices	Time (sec.)	86	434	1169	4256	13989	32954
	SAT$	-30426	-30484	-26876	-23125	-23034	-31368
	TN$	10029	10015	9997	9801.6	9804	10172
	PDN$	3823	4121	4118	4118	4119	4110
	OPT$	44278	44621	40992	37044	36975	45651
Stage Variables	Binary	256					
	Continuous	21 (Stage 1), 1621 (Stage 2)					

Table 2.4 reports the running time, satisfaction value (SAT$), construction cost (TN$), expansion cost (PDN$), and optimal value (OPT$) for the varied numbers of random scenarios. Note that, in the warm start process, the optimal solution of an antecedent with a smaller number of scenarios is used as the initial value for its succedent with a larger number of scenarios. The running time that we report refers to the entire computation time of the warm start process. The results show that the case using the EMA and IEEE 123-bus has a much larger computational cost, and for each case, the running time increases linearly at the growth of scenarios. Besides, based on the SAA method, we observe the asymptotic convergence in the case using the SiouxFalls and IEEE 33-bus. The values of SAT$, TN$, PDN$, and OPT$ meet as the number of scenarios increases. However, the large-sized case requires more scenarios to converge than that given in Table 2.4.

2.3.3 Charging Preference Analysis

The preference function $U_{ij}(\omega)$ in (2.5) characterizes drivers' charging behavior. In this study, we choose its parameters, $b(\omega)=0.6$ and $c(\omega)=0.024$, during both the peak or non-peak hours. The relationship $b(\omega)=25c(\omega)$, which we establish according to taxi drivers' average hourly wage and current charging price in New York city Indeed [2023], NYC TAXI LIMOUSINE COMMISSION [2013], means that an increase of $\$1$ per kWh in the charging cost is equivalent to the monetary value of a driver's dissatisfaction incurred from an additional 25 -minute commute to a charging station. We conduct sensitivity analysis to test the marginal impact of $b(\omega)$ and $c(\omega)$ on the optimal solution. It shows that the optimal solution is stable when $b(\omega)$ varies between 0.3 and 1.2. Increasing $c(\omega)$ substantially amplifies the influence of travel times on ET drivers' choices of charging stations, leading to a modification in the optimal decision-making

process for deploying charging stations. In contrast, decreasing $c(\omega)$ slightly does not affect the optimal solution. However, when $c(\omega)$ is reduced to 0.1, drivers become indifferent to travel distance. This, in turn, encourages a reduction in the number of charging stations and the construction of larger stations uptown.

2.3.4 Economic Performance Analysis

The three cases - Human-PDNcost, Human-TNcost, and Auto-Level - are designed to evaluate the roles of the TN and PDN as well as the benefit of utilizing autonomous ETs, compared to the baseline with $\alpha = \beta = 1$ and $\Delta U = 0$. Recall that α and β scale the costs of constructing charging stations and expanding the PDN. Human-PDNcost evaluates the role of the PDN by analyzing the marginal effect of the expansion cost at varied β, while Human-TNcost tests the role of the TN by adjusting the construction cost coefficient α. Changing ΔU from 0 (myopic human driver) to 1 (autonomous ET), Auto-Level interprets the benefit of centralized charging management and vehicular automation.

2.3.4.1 Human-PDNcost

Table 2.5: Performance in Human-PDNcost

Expansion cost coefficient (β)	Charged#	Uncharged#	CAP#	SAT$	TN$	PDN$	OPT$	ROI
0.5	365	40	668	6038	5484	2829	2275	0.72
1 (Baseline)	355	50	639	5437	5334	5193	5089	0.52
2	271	134	436	381	3935	4736	8290	0.04

Table 2.6: Strategy in Human-PDNcost

Expansion cost coefficient (β)	Station Location(Capacity)	PDN Edge(#Added Lines)	Substation Expansion(kVA)
0.5	T1(32),T5(83),T11(107),T13(62), T14(52),T15(148),T16(184)	P1-P2(+2),P2-P3(+1),P2-P19(+1),P3-P23(+2) P10-P11(+1),P23-P24(+1),P24-P25(+2)	3374
1 (Baseline)	T1(32),T5(84),T11(107),T13(38), T14(52),T15(148),T16(178)	P1-P2(+2),P2-P3(+1),P2-P19(+1) P3-P23(+2),P23-P24(+1),P24-P25(+2)	3163
2	T1(30),T5(80),T11(104), T14(47),T15(43),T16(98),T20(34)	P1-P2(+1),P2-P3(+1),P3-P23(+1)	1863

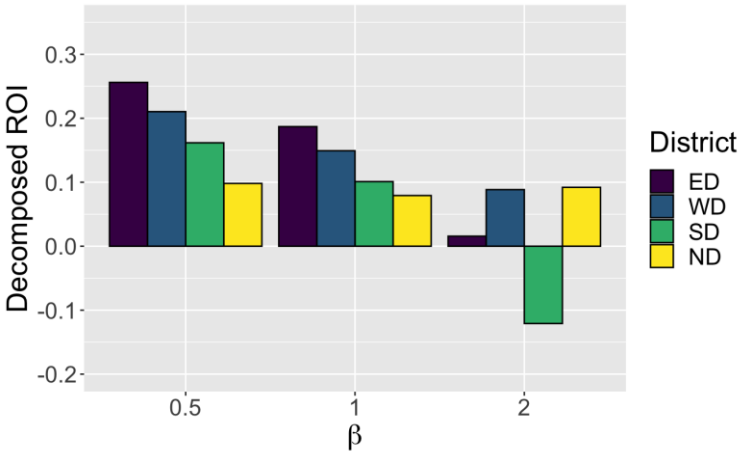
Table 2.5 displays the model's performance across different values of β. As β increases from 0.5 to 2 , notable changes occur. Specifically, there is a substantial decrease of 100 in Charged#, a reduction of 5657 in SAT$, and a decrease of 0.68 in ROI. Conversely, OPT$ increases by 6015. We also observe fluctuations in PDN$, while TN$ decreases. Notably, the proportion of PDN $ in the total consistently rises from $34.0\%\ \left(=2829/\left(2829+5484\right)\right)$ to $54.6\%\left(=4736/\left(4736+3935\right)\right)$. Figure 2.3a illustrates that the decomposed ROI of each district

decreases as β increases. Meanwhile, we find that ED and SD are sensitive to these changes. Particularly, when $\beta = 2$, SD even exhibits negative decomposed ROIs. In what follows we provide a detailed explanation of these observations, discussing the optimal solutions presented in Table 2.6. This table includes information on the locations and capacities of constructed charging stations, the number of added distribution lines, and the expansion of the substation.

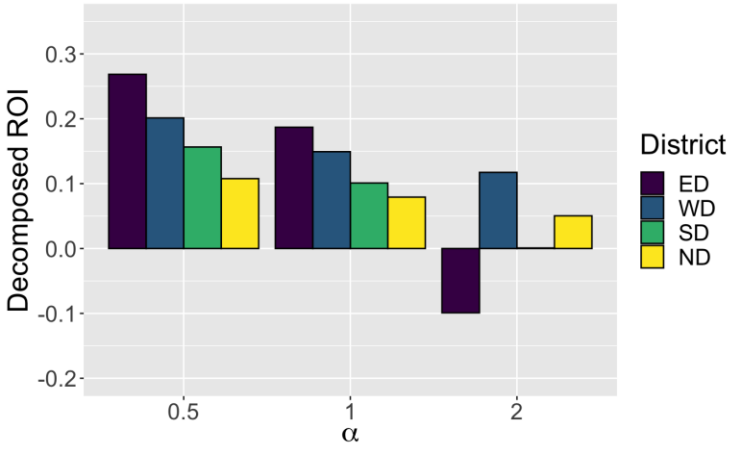
As β increases, the model is inclined to extremely reduce the extent of expanding the PDN. The capacity of the substation is reduced by 1300 and the number of distribution lines added is reduced by 7 as β increases from 0.5 to 2 . When $\beta = 0.5, 8$ distribution lines are added onto P2-P22, P2-P25 and P3-P18 to support ED, WD, and SD. When changing to the high expansion cost with $\beta = 2$, the distribution lines supporting ED and SD are canceled. In addition, 4 of the 6 distribution lines added for WD are removed. Subsequently, the charging capacity is reduced by 46.7% in ED, 36.8% at in WD, 45.2% in SD, and 4.3% in ND. Since ED has a low electricity usage and is close to the substation (see Figure 2.2b), even if the expansion distribution line (P2-P19) for ED is canceled, the original PDN is still able to support most charging demands. As a result, the decomposed ROI in ED remains positive. On the other hand, the capacity reductions in both SD and its neighbor WD greatly weaken the charging satisfaction in SD. Accordingly, the decomposed ROI in SD becomes negative.

2.3.4.2 Human-TNcost

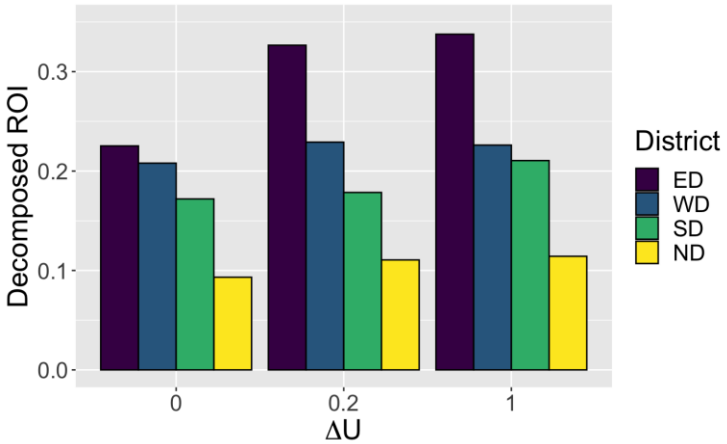
We now explore the impact of construction costs by adjusting the parameter α . Table 2.7 illustrates how varying α from 0.5 to 2 affects our model's performance. As α increases, we observe the following changes: Charged# decreases by 85, SAT$ decreases



(a) Human-PDNcost



(b) Human-TNcost



(c) Auto-Level

Figure 2.3: Decomposed ROI

Table 2.7: Performance in Human-TNcost

Construction cost coefficient (α)	Charged#	Uncharged#	CAP#	SAT$	TN$	PDN$	OPT$	ROI
0.5	370	34	717	6323	2895	5720	2292	0.73
1 (Baseline)	355	50	639	5437	5334	5193	5089	0.52
2	285	120	487	581	4663	3687	7768	0.07

Table 2.8: Strategy in Human-TNcost

Construction cost coefficient (α)	TN Location(Capacity)	PDN Edge(#Added Lines)	Substation Expansion(kVA)
0.5	T1(32), T4(83), T11(108), T13(64), T14(54), T15(148), T16(207), T20(21)	P1-P2(+2), P2-P3(+1), P2-P19(+1), P3-P23(+2) P10-P11(+1), P23-P24(+1), P24-P25(+2)	3452
1 (Baseline)	T1(32), T5(84), T11(107), T13(38), T14(52), T15(148), T16(178)	P1-P2(+2), P2-P3(+1), P2-P19(+1) P3-P23(+2), P23-P24(+1), P24-P25(+2)	3163
2	T1(22), T5(82), T11(97), T13(27), T14(44), T15(146)	P1-P2(+1), P2-P3(+1), P3-P23(+2) P23-P24(+1), P24-P25(+2)	2014

by $5754, TN\$$ increases by 1768 , and PDN $\$$ decreases by 2033 . Consequently, OPT $\$$ experiences a 5754 increase while ROI decreases by 0.66 . Figure 2.3b presents the effect of increasing α on the decomposed ROIs of the districts, with a notable impact on ED. Interestingly, this result is in contrast to the effect of adjusting β on Human-PDNcost.

As α increases, the construction cost in ED (Central Business District) becomes significantly higher compared to other districts. This prompts the removal of charging station T16 in ED, as shown in Table 2.8, resulting in a 100% reduction in charging capacity in ED. In contrast, WD experiences a 6.8% reduction, SD a 68.2% reduction, and ND a 4.4% reduction in charging capacity. Clearly, ED is much more sensitive to changes in α compared to SD and other districts.

2.3.4.3 Auto-Level

In this analysis, we examine the impact of centralized charging management and vehicular automation, denoted as ΔU . We consider three levels: human driving $(\Delta U = 0)$, semi-autonomous driving $(\Delta U = 0.2)$, and fully autonomous driving $(\Delta U = 1)$. As ΔU increases from 0 to 1 , as shown in Table 2.9, we observe the following changes: Charged# increases by 13, SAT $\$$ increases by 710 , while $TN\$$ decreases by 757 and

Table 2.9: Performance in Auto-Level

Level of vehicular automation (ΔU)	Charged#	Uncharged#	CAP#	SAT$	TN$	PDN$	OPT$	ROI
0 (Baseline)	355	50	639	5437	5334	5193	5089	0.52
0.2	379	26	619	6766	5121	5155	3510	0.66

1	368	37	578	6147	4577	4655	3085	0.67

Table 2.10: Strategy in Auto-Level

Level of vehicular automation (ΔU)	TN Location(Capacity)	PDN Edge(#Added Lines)	Substation Expansion(kVA)
0 (Baseline)	T1(32),T5(84),T11(107),T13(38), T14(52),T15(148),T16(178)	P1-P2(+2),P2-P3(+1),P2-P19(+1) P3-P23(+2),P23-P24(+1),P24-P25(+2)	3163
0.2	T1(32),T5(83),T11(142),T13(37), T14(68),T15(95),T16(162)	P1-P2(+2),P2-P3(+1),P2-P19(+1), P3-P23(+2),P23-P24(+1),P24-P25(+2)	3496
1	T1(62),T5(84),T11(136),T13(26), T14(76),T15(95),T16(98),T20(12)	P1-P2(+2),P2-P3(+1), P3-P23(+2),P23-P24(+1),P24-P25(+2)	3243

PDN $ by 538. Consequently, OPT $ decreases by 2004, but ROI increases by 0.15. Figure 2.3c further illustrates that the decomposed ROIs of all districts improve. This suggests that utilizing autonomous ETs enhances investment efficiency.

Table 2.10 reveals that as ΔU increases from 0 to 0.2 , the charging capacity decreases by 9.0% in ED, 0.7% in WD, 2.6% in SD, and 0.8% in ND. With a further increase to 1 , the charging capacity decreases by 44.9% in ED, while increasing by 28.9% in SD and 25.9% in ND.

The use of autonomous ETs allows for greater flexibility and efficiency in demand allocation to charging stations. Autonomous ETs in ED are more willing to travel longer distances to neighboring uptown districts for charging, enabling the reduction of charging station size in ED to lower costs. Additionally, a single charging station at $\mathrm{T13}$ in SD, when $\Delta U = 0$ or 0.2 , is split into two smaller stations at $\mathrm{T13}$ and T20 when $\Delta U = 1$. Smaller charging stations can lead to intense competition among myopic human drivers. Therefore, this strategy for $\Delta U = 1$ is less suitable for smaller ΔU values.

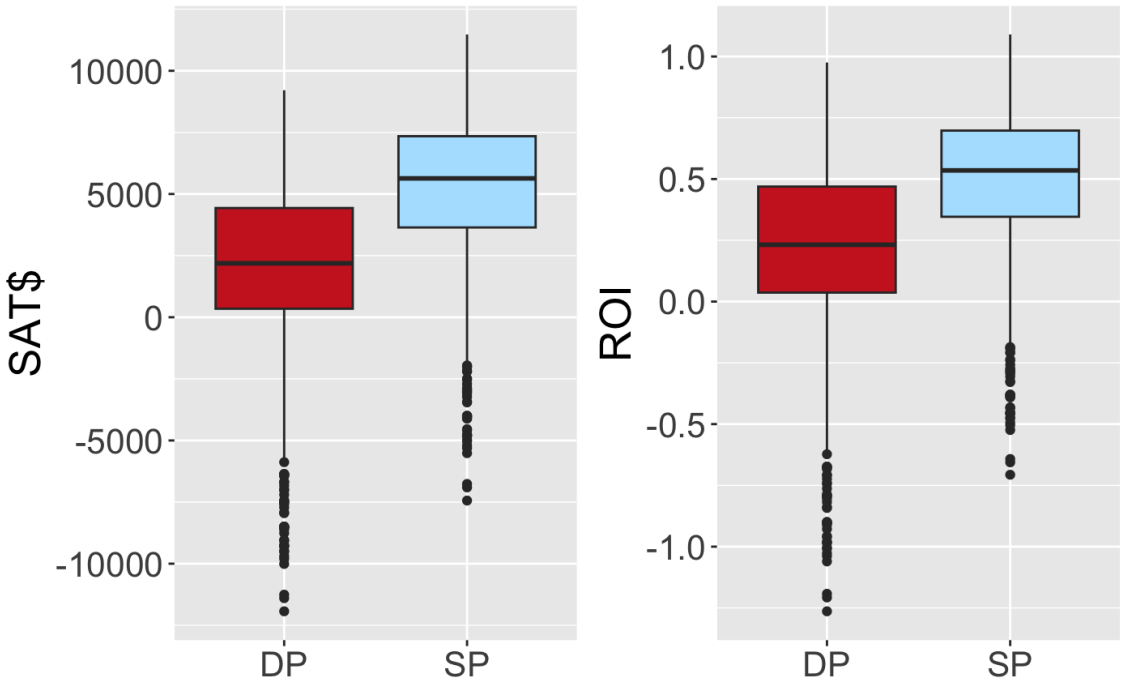


Figure 2.4: Evaluation of Stochastic Programming

2.3.5 Uncertainty Analysis

We now show the advantages of stochastic programming used in our model (called SP in short) to handle uncertainty. In the context of comparisons, we formulate a deterministic optimization model (called DP) replacing all random parameters $\left(D_i(\omega), P_n(\omega), Q_n(\omega), U_{ij}(\omega)\right)$ in the second stage of our model with their expected values. This experiment generates 108 scenarios for model training to solve optimal charging station deployment strategies and 2700 scenarios for testing to evaluate the economic performance of the strategies. The strategy recommended by SP costs $\mathrm{TN\$} = 5334$ and $\mathrm{PDN\$} = 5193$, in comparison with the lower costs of DP as TN $\$ = 4703$ and $\mathrm{PDN\$} = 4738$ (note that $\mathrm{TN\$}$ and PDN$ are fixed without regard to test scenarios). However, the higher investment yields a large improvement on the entire distribution of SAT$ (including the average and worst cases), shown in Figure 2.4 which are the box-plots of the results for all the test scenarios. Importantly, SP has a higher ROI than DP, which means that the investment according to the SP strategy is more valuable.

2.4 Conclusions

The success of taxi fleet electrification is highly reliant on adequate charging infrastructure that connects the TN and PDN. Heavily charging ET fleet on road challenges the existing TN and PDN. In this paper, we have addressed a public charging station planning problem involving the interdependence of the TN and PDN. Using two-stage stochastic programming techniques, we have presented the best planning strategy of upgrading the PDN and deploying charging stations and sought a long-run optimal policy to serve ET charging demands under the uncertainties in charging demand, traffic conditions, electricity consumption, as well as electricity price.

Distinct Marginal Effects of TN and PDN: our analysis has demonstrated that the TN and PDN have distinct marginal effects on the deployment of charging stations. This suggests that the capacity and structure of these networks can impact the placement and functioning of ET charging infrastructure.

Economic Factors Considered: Two economic factors were considered in our experimental tests. One is the cost associated with expanding the PDN, which includes increasing the capacity of transmission substations and adding power distribution lines. The other factor is the expense related to deploying charging stations, which encompasses the costs of acquiring land in TN and setting up charging stations with their associated capacities.

Conflicting Policies: These economic factors result in conflicting policies for where to construct charging stations and what their capacities should be. This implies that there are trade-offs to be considered when making decisions about charging station deployment, as increasing one factor may come at the expense of the other.

Impact on ET Drivers' Satisfaction: Regardless of the specific policies, an increase in either of these costs leads to a significant reduction in ET drivers' satisfaction with the charging process. This highlights the importance of cost management in the context of electric vehicle charging infrastructure.

Advantages of an Autonomous ET Fleet: Our analysis suggests that utilizing an autonomous ET fleet has several advantages. First, it consistently results in a higher return on investment (ROI) compared to a human-driven ET fleet. Second, an autonomous ET fleet helps in reconciling the conflicting effects of the TN and PDN on charging station deployment. This implies that autonomous vehicles might be more adaptable and efficient in utilizing charging infrastructure.

In this paper, our primary focus lies in developing a robust first-stage solution capable of effectively managing performance across a spectrum of second-stage random scenarios occurring over different time periods. Given the plethora of uncertainty factors affecting both the TN and PDN, achieving accurate results demands a substantial volume of observations. The computational complexity of our model makes it challenging to dynamically reflect ET fleet responses to changes in the infrastructure. Incorporating a game-involved second stage into our model poses a formidable computational challenge, at times rendering it unsolvable within reasonable timeframes. This complexity arises from the intricate interactions and decision-making processes inherent in game scenarios, which introduce a level of uncertainty and combinatorial complexity that can overwhelm traditional optimization methods.

Chapter 3 Managing Distributional Ambiguity in Stochastic Optimization through a Statistical Upper Bound Framework

3.1 Introduction

In both engineering and management, effective decision-making often hinges on the ability to navigate through uncertain parameters. Stochastic optimization offers a modeling framework representing the uncertain parameters as random variables, complete with their respective probabilistic information or distribution. Despite this theoretical foundation, real-world applications grapple with challenges such as limited data, incomplete information, and the inherent complexity of the systems being modeled. As a result, the exact estimation of probability distributions becomes a formidable task. This discrepancy between theoretical models and practical constraints introduces distributional ambiguity — a pervasive challenge in decision-making scenarios. This creates a substantial hurdle known as the Optimizers' Curse [Smith and Winkler, 2006], where optimizing under the influence of this ambiguity may lead to suboptimal solutions. Addressing the Optimizers' Curse becomes imperative in ensuring the practical efficacy of decision-making processes in the face of real-world complexities.

In this paper, we develop a novel approach that integrates statistical inference into stochastic optimization to address the challenges posed by distributional ambiguity. Within the field of statistical inference, frequentists tackle distributional ambiguity by considering sample uncertainty, leading to the concept of the sampling distribution of a point estimator. Utilizing repeated sampling, frequentist methods aim to comprehend the inherent variability in point

estimators. This approach facilitates the quantification of uncertainty surrounding an estimator's performance and provides a foundation for constructing upper confidence bounds within a frequentist framework.

A general expectation minimization problem can be formulated as follows:

$$\min_{x\in\mathcal{X}} \mathbb{E}_{\mathbb{P}}\left[F\left(x,\xi\right)\right], \tag{EM-M}$$

where $\mathcal{X}$ is a decision region, ξ is a vector-valued random parameter associated with probability measure $\mathbb{P}$, and F is a multivariate cost function. Suppose we lack knowledge of $\mathbb{P}$ but possess a random sample and its associated empirical distribution $\hat{\mathbb{P}}_N$. With a sufficiently large sample size N, the Sample Average Approximation (SAA) model,

$$\min_{x\in\mathcal{X}} \mathbb{E}_{\hat{\mathbb{P}}_N}\left[F\left(x,\xi\right)\right], \tag{SAA-M}$$

can serve a good estimate of EM-M. However, the estimation may exhibit significant bias when N is small. This research aims to develop a $100(1-\alpha)\%$ upper confidence bound for the expectation $\mathbb{E}_{\mathbb{P}}\left[F\left(x,\xi\right)\right]$ based on the empirical distribution $\hat{\mathbb{P}}_N$. Denote by $\mathbb{U}^{\alpha}\left[\mathbb{E}_{\mathbb{P}}\left[F\left(x,\xi\right)\right]|\ \hat{\mathbb{P}}_N\right]$ this upper confidence bound and generally describe our datadriven model as follows:

$$\min_{x\in\mathcal{X}} \mathbb{U}^{\alpha}\left[\mathbb{E}_{\mathbb{P}}\left[F\left(x,\xi\right)\right]|\ \hat{\mathbb{P}}_N\right]. \tag{UB-M}$$

UB-M incorporates a confidence bound into stochastic optimization. This innovative method effectively addresses challenges stemming from distributional ambiguity by leveraging robust statistical techniques.

In the existing literature, the recognition of distributional ambiguity highlights the necessity of deploying distributionally robust optimization (DRO) strategies (Rahimian and

Mehrotra (2019), Lin et al. (2022) and references therein). These strategies are crafted to guide decision-making processes that excel across a spectrum of plausible distributional assumptions, avoiding reliance on a single assumed distribution. Two predominant DRO approaches involve representing moment-based and discrepancybased ambiguity sets for the distributions. The moment-based approach characterizes uncertainty by imposing constraints on the moments (such as mean and variance) of the distribution (Calafiore and Ghaoui, 2006, Delage and Ye, 2010, Wiesemann et al., 2014). On the other hand, the discrepancy-based approach focuses on measuring the difference between the true distribution and a candidate distribution within the set. Examples of discrepancy-based ambiguity sets include those based on ϕ-divergence (Read and Cressie, 2012, Ben-Tal et al., 2013, Bayraksan and Love, 2015) and the Wasserstein metric (Mohajerin Esfahani and Kuhn, 2018, Blanchet and Murthy, 2019, Xie, 2020, Duque et al., 2022, Gao and Kleywegt, 2023). By embracing the inherent ambiguity associated with underlying probability distributions, DRO empowers decision-makers to formulate strategies that demonstrate resilience and effectiveness under diverse scenarios. Notably, the Wasserstein metric-based DRO has gained popularity in various fields due to its appealing properties, which include finite-sample guarantees and asymptotic consistency.

In contrast to the conventional practice of defining an ambiguity set of distributions in DRO, we propose an alternative: minimizing a confidence upper bound for the expected value of a random objective function in stochastic optimization. This approach offers a unique perspective for gauging distributional ambiguity. Inspired by robust statistical methods, we utilize the concept of an upper confidence bound for the mean. This statistical technique, specifically designed to accommodate estimation errors, provides a range of values within which the true population mean

is likely to reside. By presenting this upper bound, we not only quantify uncertainty but also enhance the reliability of our statistical inferences.

3.1.1 An Example in Charging Planning

Example 3.1. We consider a day-ahead charging model for an electric vehicle (EV) fleet where the goal is to minimize the total expected cost of charging the fleet under uncertain electricity prices and demand. The optimization model is formulated as follows:

$$\min_{x}\left\{cx+\mathbb{E}_{p}\left[Q\left(x,\xi\right)\right]:cx<B,x\geq 0\right\}, \tag{3.1}$$

where the cost function $Q\left(x,\xi\right)$ is given by

$$Q\left(x,\xi\right):=\min_{y}\left\{s\left(D\left(\xi\right)-x-y\right)+W\left(\xi\right)y:W\left(\xi\right)y+cx\leq B,y\geq 0\right\}. \tag{3.2}$$

Here, x represents the amount of eletricity to buy now and y is the amount to buy tomorrow, $D\left(\xi\right)$ signifies the demand for electricity, which is dependent on the random variable ξ, encapsulating the uncertainty in demand. The term $W\left(\xi\right)$ reflects the price of electricity that can vary with ξ. The penalty for unsatisfied demand is represented by s. The decision variables x and y denote the electricity purchased a day ahead and on the day of operation, respectively.

Numerical Test. In our study, we set the unit price of buying electricity today as $c=1$. We differentiate between two types of scenarios: regular days and worst-case days. On regular days, the distributions for charging demand $\left(D\left(\xi\right)\right)$ and price $\left(W\left(\xi\right)\right)$ are characterized as follows:

$$D\left(\xi\right)\sim\mathcal{U}\left(5,15\right),W\left(\xi\right)\sim\mathcal{U}\left(0.5,1.5\right),$$

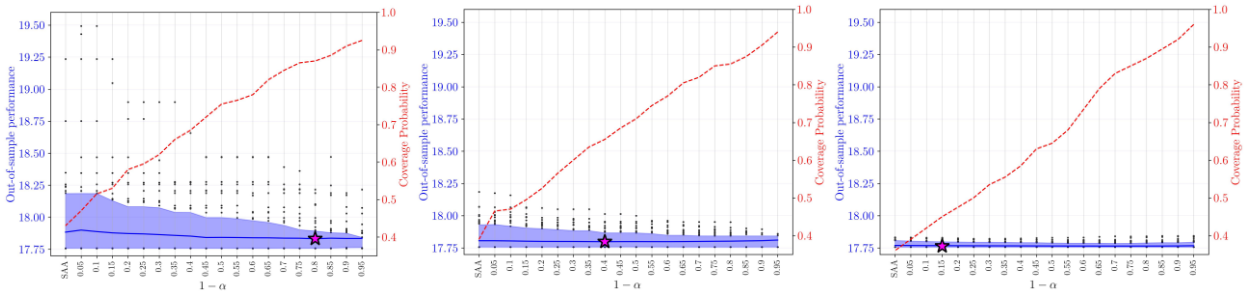
where $\mathcal{U}(a,b)$ denotes the uniform distribution between a and b. In the worst-case scenario, the mean values of these random variables are threefold compared to regular days, specifically:

$$D(\xi) \sim 2 \cdot \mathcal{U}(5,15), W(\xi) \sim 2 \cdot \mathcal{U}(0.5,1.5).$$

The probabilities of encountering regular and worst-case days are also differentiated, with P (regular $) = 0.7$ and P (worst $) = 0.3$. The total buget B is set to 12 and the penalty $s = 2.5$. Figure 3.1 demonstrates that the SAA method tends to be overly optimistic, especially when only a small sample size is available. In contrast, our approach provides a more robust solution.

3.1.2 Literature review of statistical confidence interval

One-sided and two-sided confidence intervals, a well-explored domain in statistical theory, are renowned for their robust asymptotic correctness, accuracy, and consistency, directly linked to sample size. A confidence interval is deemed first-order accurate if its confidence level error is within the inverse of the square root of the sample size, and second-order accurate if within the inverse of the sample size Vaart [1998].



(a) $N = 10$ (b) $N = 30$ (c) $N = 70$

Figure 3.1: The optimal value of our model is 29.7. We now compare the result for Out-of-sample performance (left axis, solid line, and shaded area) and the coverage probability (right axis, dashed

line) as a function of the nominal confidence level $(1-\alpha)$ in APUB-SP. The star symbol indicates the point where the mean of the out-of-sample performance attains its minimum.

The inherent convergence with the sample size is a notable feature of those upper confidence bounds. Moreover, statistical upper bounds provide additional advantages rooted in their well-established theoretical foundations. The definitions and concepts underpinning those bounds have undergone thorough scrutiny and refinement in the field of statistics. This rigorous academic exploration has yielded a comprehensive and uniform framework, facilitating a nuanced understanding and practical application of those bounds. Consequently, incorporating a statistical upper bound into stochastic optimization offers not only robustness and reliability but also interpretation of our proposed optimization framework.

The most classical frequentist asymptotic approach, which utilizes the sample mean, standard deviation, and normal approximation, has been extensively discussed (see Devore (2009) and references therein). This method is particularly favored in practice for sample sizes larger than 30 , offering a straightforward yet effective means of estimating confidence intervals Hazra (2017). It hinges on the Central Limit Theorem (CLT), which posits that the distribution of a sample mean approximates a normal distribution as the sample size increases, regardless of the population's distribution. Efron's bootstrap confidence interval Efron (1981), another method for constructing statistical upper bounds, employs the $100(1-\alpha)$-th percentile of the bootstrap distribution of a sample mean. Both these methods achieve first-order accuracy. Building upon these concepts, Efron (1987) developed bias-corrected and accelerated bootstrap confidence interval $(\mathrm{BC}_{\mathrm{a}})$ achieving a second-order accuracy. This approach fine-tunes the nominal confidence level as a function of the sample size. This advancement marks a significant step towards more precise

confidence interval estimation, especially in cases where first-order methods might not suffice due to smaller sample sizes or more complex data structures. In machine learning, the construction of upper confidence bounds often leverages concentration inequalities to ensure robustness against data variability. Hoeffding-type bounds (Auer et al., 2002) are popular for their simplicity and effectiveness in bounding the sum of bounded random variables, particularly useful in scenarios with limited prior knowledge about data distributions. Empirical Bernstein-type bounds (Mnih et al., 2008) offer improvements over Hoeffding's approach by incorporating sample variance, making them more adaptable to data with varying degrees of variability. This type of bound is particularly advantageous in dealing with heteroskedastic data, where the variance is not constant. The self-normalized bounds (Abbasi-yadkori et al., 2011), on the other hand, are designed to handle the challenges of auto-correlated data, common in time-series analysis. These bounds normalize the sum of random variables by their cumulative variance, offering a more dynamic approach to uncertainty quantification in sequential decision processes. While these bounds are grounded in robust statistical principles, they each have limitations: they may be nonconvex or not sufficiently data-driven, leading to challenges in optimization or an overly conservative nature.

3.1.3 Contributions and Organization of this Paper

The primary contributions of this paper are summarized as follows:

We introduce the Average Percentile Upper Bound (APUB), a novel statistical construct that serves as an upper bound for population means and a risk metric for sample means. The robust statistical foundations of APUB are established through rigorous proofs of its asymptotic correctness and consistency, offering a reliable basis for its applications in data-driven decision-making.

The innovative integration of APUB into stochastic optimization frameworks mitigates the ambiguity stemming from sparse data in probability distributions. The theoretical properties of APUB are adeptly applied to a new optimization framework, simultaneously ensuring model reliability and interpretability while reducing over-conservatism. This synergy narrows the gap between statistical upper bounds and stochastic optimization, fostering theoretical advancement and practical utility.

The development of a bootstrap sampling approximation method is tailored for solving APUB-embedded optimization frameworks, particularly focusing on twostage linear stochastic optimization with random recourse. We present formal proofs of the method's stability and convergence, underscoring its effectiveness and applicability.

The rest of this paper is organized as follows: Section 3.2 introduces the concept of APUB. Specifically, Section 3.2.1 formally defines APUB and explores its statistical implications, while Section 3.2.2 provides a thorough examination of the asymptotic properties of APUB, including the proofs of asymptotic correctness and consistency. Section 3.3 details the integration of APUB into stochastic optimization. Section 3.3.1 addresses the asymptotic correctness of our optimization framework, and Section 3.3.2 establishes its asymptotic consistency. Section 3.4 develops a bootstrap sampling approximation approach to solve our optimization framework. We also show the convergence and stability of the approximation. Section 3.5 presents a comprehensive numerical analysis, applying the proposed framework across a variety of classical stochastic optimization scenarios. The paper concludes with Section 3.6, summarizing key findings and contributions.

3.2 Average Percentile Upper Bound

We establish formal definitions for the key concepts employed in this paper. Consider an induced probability space $(\Xi, \mathfrak{B}, \mathbb{P})$, where Ξ is the support of a random vector, $\mathfrak{B}$ is the Borel σ-algebra, and $\mathbb{P}$ is a probability measure. Let $(\xi_1, \ldots, \xi_N) \sim \mathbb{P}$ indicate an independent and identically distributed (i.i.d) random sample with a size of N generated from $(\Xi, \mathfrak{B}, \mathbb{P})$. The empirical distribution associated with the random sample is represented as

$$\hat{\mathbb{P}}_N := \frac{1}{N}\sum_{n=1}^{N} \delta_{\xi_n},$$

where δ_{ξ_n} is the Dirac delta function at ξ_n. As N increases to infinity, we have a sample path $(\xi_1, \xi_2, \ldots)$. Without loss of generality, we ignore the decision variable x and focus our discussion on a measurable cost function $F : \Xi \to \mathbb{R}$ in this section. Denote by $\mu := \mathbb{E}_{\mathbb{P}}\left[F(\xi)\right]$ the population mean and by $\sigma^2 := \mathbb{E}_{\mathbb{P}}\left[(F(\xi) - \mu)^2\right]$ the population variance. We assume μ and σ to be finite in the late statement. Also let $\hat{\mu}_N := \mathbb{E}_{\hat{\mathbb{P}}_N}\left[F(\xi)\right]$ be the sample mean and $\hat{S}_N := \mathbb{E}_{\hat{\mathbb{P}}_N}\left[F(\xi) - \hat{\mu}_N\right]$ be the sample variance.

3.2.1 Concept of Average Percentile Upper Bound

Using the bootstrap percentile method, Efron [1981] presents a $100(1-\alpha)\%$ bootstrap based upper confidence bound for μ as

$$U_M^{\alpha} := \inf\left\{t \in \mathbb{R} : \Pr\left(\mu^* \le t \mid \mathbb{P}_*\right) \ge 1-\alpha\right\}. \tag{3.3}$$

where $\mathbb{P}_*$ is a bootstrap distribution and μ^* is a bootstrap estimator of μ. Equation (3.3) is explicitly represented as the limit of

$$U_M^\alpha = \inf\left\{t \in \mathbb{R} : \frac{1}{M}\sum_{m=1}^{M} 1\left\{\frac{1}{N}\sum_{n=1}^{N} F\left(\zeta_{m,n}\right) \le t\right\} \ge 1-\alpha\right\}, \tag{3.4}$$

where $1\{\cdot\}$ is the indicator function and $\left(\zeta_{m,1},\ldots,\zeta_{m,N}\right) \sim \hat{\mathbb{P}}_N$, for $m = 1,\ldots,M$, are bootstrap samples. The following proposition depicts the limit of U_M^α.

Proposition 3.2. For a given $\hat{\mathbb{P}}_N$, we let

$$\mathbb{U}_{\text{Efron}}^\alpha\left[\mu \middle| \hat{\mathbb{P}}_N\right] := \inf\left\{t \in \mathbb{R} : \Pr\left(\frac{1}{N}\sum_{n=1}^{N} F\left(\zeta_n\right) \le t \middle| \hat{\mathbb{P}}_N\right) \ge 1-\alpha\right\}, \tag{3.5}$$

where $\left(\zeta_1,\cdots,\zeta_N\right) \sim \hat{\mathbb{P}}_N$. Then, as $M \to \infty, U_M^\alpha \to \mathbb{U}_{\text{Efron?}}^\alpha\left[\mu \middle| \hat{\mathbb{P}}_N\right]$ w.p. 1 (for ζ).

Proof. For a fixed $\hat{\mathbb{P}}_N$, we briefly denote by $\boldsymbol{\zeta} := \left(\zeta_1,\ldots,\zeta_N\right)$ a random matrix whose probability measure is an N -fold Cartesian product $\hat{\mathbb{P}}_N \times \cdots \times \hat{\mathbb{P}}_N$. Thus $\boldsymbol{\zeta}_m := \left(\zeta_{m,1},\ldots,\zeta_{m,N}\right)$, for $m = 1,\ldots,M$, is a random sample from $\hat{\mathbb{P}}_N \times \cdots \times \hat{\mathbb{P}}_N$. Let $G\left(\boldsymbol{\zeta}\right) := \frac{1}{N}\sum_{n=1}^{N} F\left(\zeta_n\right)$. It means that $\Psi\left(t\right) := \Pr\left(G\left(\boldsymbol{\zeta}\right) \le t \middle| \hat{\mathbb{P}}_N\right)$ is the cumulative probability distribution (cdf) of $G\left(\boldsymbol{\zeta}\right)$ and $\Psi_M\left(t\right) := \frac{1}{M}\sum_{m=1}^{M} 1\left\{G\left(\boldsymbol{\zeta}_m\right) \le t\right\}$ is an empirical cdf associated with a random sample $G\left(\boldsymbol{\zeta}_1\right),\ldots,G\left(\boldsymbol{\zeta}_M\right)$. Thus, $\Psi_M\left(t\right)$ uniformly converges to $\Psi\left(t\right)$ w.p. 1 as M goes to infinity. Therefore, any percentile of $\Psi_M\left(t\right)$ converges to the counterpart of $\Psi\left(t\right)$ w.p.1. It follows that U_M^α converges to $\mathbb{U}_{\text{Efron}}^\alpha\left[\mu \middle| \hat{\mathbb{P}}_N\right]$ w.p.1.

While $\mathbb{U}^{\alpha}_{\text{Efron?}}\left[\mu|\ \hat{\mathbb{P}}_N\right]$, the percentile-based upper bound, is satisfactory in many statistical analyses, it is non-convex and difficult to control/optimize for highly skewed distributions, which are regarded as inferior properties in the realm of optimization. Therefore, we extend Efron's upper bound by averaging over the values to the right of the $100(1-\alpha)$-th percentile.

Definition 3.3. The average percentile upper bound for μ with a nominal confidence level $(1-\alpha)$ is denoted as

$$\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu|\ \hat{\mathbb{P}}_N\right] := \frac{1}{\alpha}\int_0^{\alpha}\mathbb{U}^{\tau}_{\text{Efron}}\left[\mu|\ \hat{\mathbb{P}}_N\right]d\tau. \qquad \text{(APUB)}$$

We can interpret Efron's upper bound, alternatively in the realm of risk management and decision making, as an approximation of the Value at Risk (VaR) of $\hat{\mu}_N$ by substituting $\hat{\mathbb{P}}_N$ for $\mathbb{P}$ in the following VaR equation:

$$\text{VaR}_{\alpha}\left(\hat{\mu}_N\right) = \inf\left\{t\in\mathbb{R}:\Pr\left(\frac{1}{N}\sum_{n=1}^{N}F\left(\xi_n\right)<t\middle|\mathbb{P}\right)>1-\alpha\right\}.$$

Analogously, APUB approximates the Conditional Value at Risk (CVaR) of $\hat{\mu}_N$,

$$\text{CVaR}_{\alpha}\left(\hat{\mu}_N\right) = \frac{1}{\alpha}\int_0^{\alpha}\text{VaR}_{\tau}\left(\hat{\mu}_N\right)d\tau.$$

APUB serves a dual purpose: as an upper bound for the population mean in statistics and as an approximate risk measure for the sample mean in risk assessment. As a risk measure, it primarily focuses on approximating the tail distribution of the potential estimation error of the population mean, which could result from an inadequacy of sample points. Furthermore, APUB complies with fundamental properties of a coherent risk measure, such as sub-additivity, homogeneity, convexity, translational invariance, and monotonicity. These characteristics make APUB a good candidate to be applied to stochastic optimization under distributional ambiguity, particularly in scenarios

requiring solvability, such as two-stage stochastic optimization with random recourse. Analogous to Theorem A. 2 [Rockafellar and Uryasev, 2000], the following proposition provides an alternative representation for APUB.

Proposition 3.4.

$$\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu|\ \hat{\mathbb{P}}_N\right]=\inf_{t\in\mathbb{R}}\left\{t+\frac{1}{\alpha}\int\left[\frac{1}{N}\sum_{n=1}^{N}F\left(\zeta_n\right)-t\right]\prod_{n=1}^{N}\hat{\mathbb{P}}_N\left(d\zeta_n\right)\right\}, \tag{3.6}$$

where the bold integral symbol means an N-fold integral over the N-fold Cartesian product of $\hat{\mathbb{P}}_N$.

Remark 3.5. By Proposition 3.4, we have that $\mathbb{U}^{\alpha}_{\text{APUB?}}\left[\mu|\ \hat{\mathbb{P}}_N\right]$ monotonically decreases in $\alpha\in(0,1]$ w.p.1. This implies that, for $\alpha\in(0,1]$,

$$\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu|\ \hat{\mathbb{P}}_N\right]\geq\mathbb{U}^{1}_{\text{APUB}}\left[\mu|\ \hat{\mathbb{P}}_N\right]=\int\left[\frac{1}{N}\sum_{n=1}^{N}F\left(\zeta_n\right)\right]\prod_{n=1}^{N}\hat{\mathbb{P}}_N\left(d\zeta_n\right)=\hat{\mu}_N,\text{ w.p.1.}$$

The quality of a statistical upper bound refers to the rate of its true coverage probability increasing beyond the nominal confidence level $(1-\alpha)$ as the sample size grows. Example 3.6 illustrates the following two attractive asymptotic characteristics of APUB. A theoretical discussion is given in Section 3.2.2.

1.*Asymptotic Correctness* (defined in Vaart [1998, Section 23.3]): A statistical upper bound $\mathbb{U}^{\alpha}\left[\mu|\ \hat{\mathbb{P}}_N\right]$ for μ is correct at level $(1-\alpha)$ up to κ thorder if its coverage probability

$$\Pr\left(\mu\leq\mathbb{U}^{\alpha}\left[\mu|\ \hat{\mathbb{P}}_N\right]|\ \mathbb{P}\right)\geq(1-\alpha)+O\left(N^{-\kappa/2}\right)$$

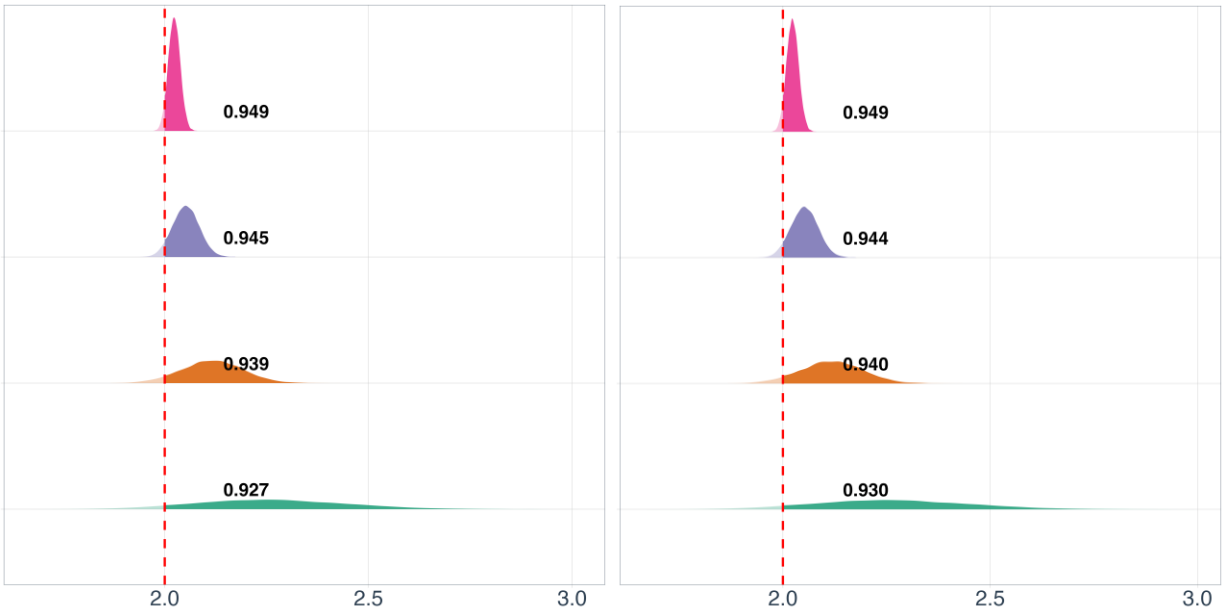
If the equality holds, one says that $\mathbb{U}^{\alpha}\left[\mu|\ \hat{\mathbb{P}}_N\right]$ is κ th-order accurate Hall [1986]. Efron's upper bound is first-order accurate and APUB is first-order correct. Notably, the terms 'asymptotic

correctness' and 'asymptotic accuracy' are different. Specifically, the asymptotic correctness implies that, when the sample size N is sufficiently large, the nominal confidence level serves as a reliable lower bound for the coverage probability.

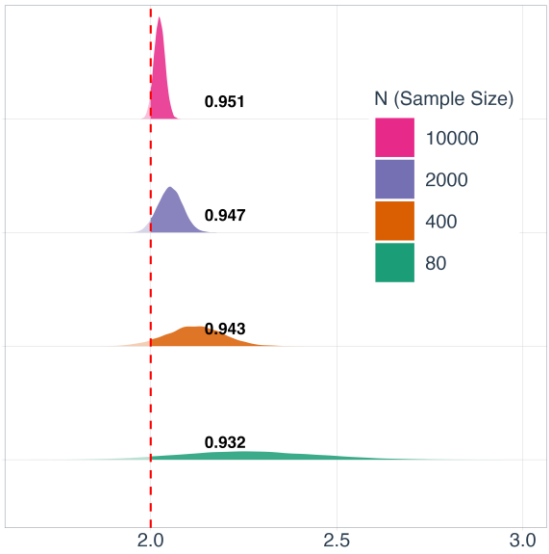
2.*Asymptotic consistency*: APUB converges to the population mean w.p. 1 as the sample size increases to infinity. This attribute ensures that APUB is a consistent estimator for the population mean.

Example 3.6. Let $\xi \sim \text{Gamma}(2,1)$ and $F(\xi) = \xi$. So the population mean $\mu = 2$. We compare APUB with Efron's upper bound and the standard large-sample upper bound given as $\hat{\mu}_N + z_\alpha S_N / \sqrt{N}$, where z_α denotes z critical value. In order to estimate the probability density functions (pdf) of three upper bounds, we performed a Monte Carlo simulation with $\alpha = 0.05$ while allowing the sample sizes, N, to vary from 80 to 10,000 .

As illustrated in Figure 3.2, the coverage probability is essentially the area to the right of the vertical dotted line at $\mu = 2$ in a pdf. Our results show that as N increases, the coverage probabilities for both the large-sample and Efron's upper bounds get closer to $(1-\alpha) = 0.95$. This demonstrates the asymptotic accuracy of these two types of bounds. However, this is in stark contrast to APUB which doesn't have asymptotic accuracy. In fact, as N approaches infinity, the coverage probability of APUB can



(a) Standard Upper Bound (b) Efron's Upper Bound



(c) APUB

Figure 3.2: The comparison between APUB, Efron's upper bound, and the standard large-sample upper bound.

grow beyond $(1-\alpha)$. Moreover, this growth rate is observed to be more rapid than the other two bounds, which brings attention to the unique nature of APUB.

Furthermore, all three upper bounds exhibit asymptotic consistency. As N increases, they all converge to $\mu = 2$ w.p.1. This essentially means they become more precise as more data is collected. By examining the pdf curves, it is apparent that they narrow and concentrate more intensely around μ, which visually indicates this trend.

3.2.2 Asymptotic Characteristics of APUB

We now theoretically discuss the asymptotic correctness and consistency of APUB. The following proposition shows its asymptotic correctness.

Proposition 3.7. Suppose that the skewness $\mathbb{E}_{\mathbb{P}}[F(\xi)-\mu]^3/\sigma^3 < \infty$. Then, for a fixed nominal confidence level $1-\alpha, \mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu|\ \hat{\mathbb{P}}_N\right]$ is 1st-order asymptotically correct, i.e.,

$$\Pr\left(\mu \le \mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu|\ \hat{\mathbb{P}}_N\right]|\ \mathbb{P}\right) \ge (1-\alpha) + O\left(N^{-1/2}\right)$$

Proof. Theorem A. 3[Efron, 1981] shows that $\mathbb{U}^{\alpha}_{\text{Efron?}}\left[\mu|\ \hat{\mathbb{P}}_N\right] = (1-\alpha) + O\left(N^{-1/2}\right)$. By Definition 3.3, we know that $\mathbb{U}^{\alpha}_{\text{APUB?}}\left[\mu|\ \hat{\mathbb{P}}_N\right] \ge \mathbb{U}^{\alpha}_{\text{Efron?}}\left[\mu|\ \hat{\mathbb{P}}_N\right]$.

Proposition 3.7 establishes the asymptotic correctness of APUB, which guarantees that the nominal confidence level is a conservative boundary for the actual coverage probability. This attribute confirms that APUB is an effective upper-bound statistic, especially valuable for its robust response to distributional ambiguity encountered with limited sample data. Considering the uncertainty diminishes along with an increase in the sample size, we next show that APUB is a consistent estimator for the population mean.

Theorem 3.8. For any $\alpha \in (0,1]$, as $N \to \infty$,

$$\mathbb{U}^{\alpha}_{APUB}\left[\mu \middle| \hat{\mathbb{P}}_N\right] \to \mu, w.p.1.$$

To prove Theorem 3.8, we need the following lemma about the bootstrap law of large numbers.

Lemma 3.9. Let $(\zeta_1,\ldots,\zeta_N) \sim \hat{\mathbb{P}}_N$. Then, as $N \to \infty$,

$$\frac{1}{N}\sum_{n=1}^{N} F(\zeta_n) \to \mu w.p.1.$$

Proof. According to Theorem A. 4 [Athreya, 1983], if $\liminf_{M,N\to\infty} MN^{-\phi} > 0$ for some $\phi > 0$, and $\mathbb{E}_{\mathbb{P}} |F(\xi) - \mu|^{\theta} < \infty$ for some $\theta \geq 1$ such that $\theta\phi > 1$, we have that, as $M, N \to \infty$,

$$\frac{1}{M}\sum_{m=1}^{M} F(\zeta_m) \to 1\text{w.p.}1$$

where $(\zeta_1,\ldots,\zeta_M) \sim \hat{\mathbb{P}}_N$

In our case, choose $\phi = 1, \theta = 2$, and $M = N$. This ensures that $\liminf_{M,N\to\infty} MN^{-\phi} = 1 > 0$. The condition $\mathbb{E}_{\mathbb{P}} |Z(\xi) - \mu|^{\theta} < \infty$ is satisfied due to finite variance. This completes the proof.

Proof of Theorem 3.8. Let $(\bar{\xi}_1, \bar{\xi}_2, \ldots)$ be a realization of the sample path and $\bar{\mathbb{P}}_N$ be the empirical distribution associated with the first N sample points. Denote by $(\zeta_1(\bar{\mathbb{P}}_N),\ldots,\zeta_N(\bar{\mathbb{P}}_N)) \sim \bar{\mathbb{P}}_N$ a random sample under $\bar{\mathbb{P}}_N$. Denote a collection of realizations as

$$\mathfrak{S} := \left\{ (\bar{\xi}_1, \bar{\xi}_2, \ldots) : \begin{array}{ll} \lim\limits_{N\to\infty} & \frac{1}{N}\sum_{n=1}^{N} F(\bar{\xi}_n) = \mu \\ \lim\limits_{N\to\infty} & \frac{1}{N}\sum_{n=1}^{N} F(\zeta_n(\bar{\mathbb{P}}_N)) = \mu ? \text{w.p.}1 (\text{for} \zeta) \end{array} \right\}. \tag{3.7}$$

It follows by the Strong Law of Large Number and Lemma 3.9 that $\Pr\left\{\left(\xi_1,\xi_2,\ldots\right)\in \mathfrak{S}|\ \mathbb{P}\right\}=1$.

We now fix $\left(\bar{\xi}_1,\bar{\xi}_2,\ldots\right)\in\mathfrak{S}$ along with its corresponding $\left(\bar{\mathbb{P}}_1,\bar{\mathbb{P}}_2,\ldots\right)$. Here, $\left(\bar{\xi}_1,\bar{\xi}_2,\ldots\right)$ and $\bar{\mathbb{P}}_N$ are deterministic, while $\left(\zeta_1\left(\bar{\mathbb{P}}_N\right),\ldots,\zeta_N\left(\bar{\mathbb{P}}_N\right)\right)$ remains randomness. For clarity, denote

$$\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right):=\frac{1}{N}\sum_{n=1}^{N}F\left(\zeta_n\left(\bar{\mathbb{P}}_N\right)\right).$$

By Proposition 3.4, we have

$$\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu|\ \bar{\mathbb{P}}_N\right]=\text{CVaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right).$$

Then, to prove Theorem 3.8, it suffices to show

$$\lim_{N\to\infty}\text{CVaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)=\mu. \tag{3.8}$$

Sarykalin et al. [2008] provide an expression of CVaR as

$$\text{CVaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)=\eta_{\alpha}\text{VaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)+\left(1-\eta_{\alpha}\right)\text{CVaR}^{+}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right), \tag{3.9}$$

where

$$\text{CVaR}^{+}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)=\mathbb{E}\left[\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)|\ \hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)>\text{VaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)\right],$$

$$\eta_{\alpha}=\frac{\Pr\left\{\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\le\text{VaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)\right\}-\alpha}{1-\alpha}.$$

Describe $\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)>\text{VaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)$ as event $\mathcal{A}_N$. By the definition of VaR, it is clear to see $\Pr\left(\mathcal{A}_N\right)\ge\alpha$ for all N. The expression (3.9) implies that

$$\text{VaR}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)\le\text{CVaR}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right)\le\text{CVaR}^{+}_{\alpha}\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right).$$

According to Theorem A. 5 [Vaart, 1998], we understand that convergence in distribution implies the convergence of the quantile function. Thus, we have

$$\lim_{N\to\infty} \mathrm{VaR}_\alpha\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right) = \mu,$$

On the other hand, applying Theorem A. 6 [Mallows and Richter, 1969], we have

$$\begin{aligned}\left|\mathrm{CVaR}_\alpha^+\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right) - \mathbb{E}\left[\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right]\right| &= \left|\mathbb{E}\left[\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)|\ \mathcal{A}_N\right] - \mathbb{E}\left[\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right]\right| \\ &\le \tilde{\sigma}_N\left(\frac{1-\Pr\left(\mathcal{A}_N\right)}{\Pr\left(\mathcal{A}_N\right)}\right)^{1/2} \\ &\le \tilde{\sigma}_N\left(\frac{1-\alpha}{\alpha}\right)^{1/2},\end{aligned}$$

where $\tilde{\sigma}_N$ represents the standard deviation of $\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)$. Since $\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)$ converges to μ w.p.1, we know that $\tilde{\sigma}_N$ converges to 0. Therefore, we establish:

$$\begin{aligned}\lim_{N\to\infty}\mathrm{CVaR}_\alpha^+\left(\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right) &= \lim_{N\to\infty}\mathbb{E}\left[\hat{\mu}_N\left(\bar{\mathbb{P}}_N\right)\right] \\ &\overset{(a)}{=} \lim_{N\to\infty}\mathbb{E}\left[F\left(\zeta_1\left(\bar{\mathbb{P}}_N\right)\right)\right] \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}F\left(\bar{\xi}_n\right) \\ &\overset{(b)}{=} \mu,\end{aligned}$$

where (a) holds since $\left(\zeta_1\left(\bar{\mathbb{P}}_N\right),\ldots,\zeta_N\left(\bar{\mathbb{P}}_N\right)\right)$ are drawn i.i.d. from $\bar{\mathbb{P}}_N$, and (b) holds due to the definition of $\mathfrak{S}$ (see (3.7)). Hence, we obtain (3.8) and complete the proof.

3.3 Optimization with APUB

In this section we apply APUB to stochastic optimization problems. In the context of optimization, let the cost function $F\left(x,\xi\right):\mathcal{X}\times\Xi\mapsto\mathbb{R}$ be $\mathfrak{B}$-measurable for all $x\in\mathcal{X}$. Denote

the mean and standard deviation of $F(x,\xi)$ by $\mu(x)$ and $\sigma(x)$ respectively. The UB-M framework using APUB is written as

$$\hat{\vartheta}_N^{\alpha} = \min_{x\in\mathcal{X}} \mathbb{U}_{\mathrm{APUB}}^{\alpha}\left[\mu(x)|\ \hat{\mathbb{P}}_N\right]. \tag{APUB-M}$$

By Remark 3.5, we know that $\hat{\vartheta}_N^{\alpha}$ decreases in $\alpha\in(0,1]$ w.p. 1 and $\hat{\vartheta}_N^{1}$ is the optimal value of SAA-M. Let $\hat{\mathcal{S}}_N^{\alpha}$ denote the set of optimal solutions to APUB-M. Also, denote by ϑ^* the optimal value of EM-M and by $\mathcal{S}$ the set of its optimal solutions. We now present some mild assumptions as follows.

Assumption 3.10. There exists a compact set $\mathcal{K}\subseteq\mathcal{X}$ such that:

(A1) $\mathcal{S}\subseteq\mathcal{K}$;

(A2) $\hat{\mathcal{S}}_N^{\alpha}\subseteq\mathcal{K}$ w.p. 1 for sufficiently large N and $\alpha\in(0,1]$.

Assumption 3.10 is frequently encountered in the literature pertaining to the asymptotic analysis of the SAA method [Birge and Louveaux, 2011, Shapiro et al., 2021]. This assumption posits that it is adequate to confine the examination of decision properties to the compact set $\mathcal{K}$. For the purposes of the discussion in the remainder of Section 3.3, we proceed under the premise that the decision space is indeed $\mathcal{K}$, a simplification that does not limit the generality of our analysis.

Assumption 3.11. There exists an open convex hull $\mathcal{N}$ containing $\mathcal{K}$ such that:

(B1) $F(x,\xi)$ is convex on $\mathcal{N}$ for each $\xi\in\Xi$;

(B2) $\mu(x)$ and $\sigma(x)$ are finite for all $x\in\mathcal{N}$.

Building on the aforementioned assumptions, we can draw the following observations in Proposition 3.12.

Proposition 3.12. Suppose Assumption 3.11 holds. Then $\mu(x)$ is continuous on $\mathcal{N}$, and $\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(x)|\ \hat{\mathbb{P}}_N\right]$ is a continuous convex function on $\mathcal{N}$ w.p.1.

Proof. Since $F(x,\xi)$ is convex on $\mathcal{N}, \mu(x)$ is also convex on $\mathcal{N}$. Hence, $\mu(x)$ is continuous. Under Assumption 3.11, $F(\cdot,\xi)$ is continuous and convex on $\mathcal{N}$. It is easy to see that, by Proposition 3.4, $\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(x)|\ \hat{\mathbb{P}}_N\right]$ is continuous convex on $\mathcal{N}$.

In Sections 3.3.1 and 3.3.2, we examine asymptotic characteristics of APUB-M, focusing on its data-driven nature which includes aspects such as reliability, consistency, and ease of interpretation.

3.3.1 Asymptotic Correctness

Mohajerin Esfahani and Kuhn [2018] introduce the concept of reliability for a certain optimal solution in DRO approaches. The reliability refers to the probability that the optimal value of a DRO model exceeds the expected cost of the system at the optimal solution in true scenarios. We extend this concept to the entire optimal solution set, which in our case is called the coverage probability of the general UB-M framework. Denote a probability function of a given subset $S \subseteq \mathcal{X}$ as

$$\beta(\vartheta, S) := \Pr\left(\vartheta \geq \max_{x\in S}\mu(x)|\ \mathbb{P}\right). \tag{3.10}$$

Let $\bar{\vartheta}^{\alpha}_N$ and $\bar{\mathcal{S}}^{\alpha}_N$ be the optimal value and optimal solution set of UB-M, respectively. The coverage probability of UB-M is $\beta\left(\bar{\vartheta}^{\alpha}_N, \bar{\mathcal{S}}^{\alpha}_N\right)$, which measures the chance that $\bar{\vartheta}^{\alpha}_N$ can serve as an

upper bound of the actual performance of UB-M across all optimal solutions. In the following, we define the asymptotic correctness of UB-M.

Definition 3.13. UB-M is κ th order asymptotically correct if its coverage probability converges to the nominal confidence level with a rate up to $O\left(N^{-\kappa/2}\right)$ as

$$\beta\left(\bar{\vartheta}_N^{\alpha}, \bar{\mathcal{S}}_N^{\alpha}\right) \geq (1-\alpha) + O\left(N^{-\kappa/2}\right).$$

Defined on the entire optimal solution set, the concept of asymptotic correctness is stricter than the reliability concerning a certain optimal solution. If a UB-M framework is asymptotically correct, we have that, for any $\bar{x} \in \bar{\mathcal{S}}_N^{\alpha}$,

$$\beta\left(\bar{\vartheta}_N^{\alpha}, \{\bar{x}\}\right) \geq \beta\left(\bar{\vartheta}_N^{\alpha}, \bar{\mathcal{S}}_N^{\alpha}\right) \geq (1-\alpha) + O\left(N^{-\kappa/2}\right).$$

In the subsequent statement, we refer to $\beta\left(\bar{\vartheta}_N^{\alpha}, \{\bar{x}\}\right)$ as the coverage probability of UBM concerning $\bar{x}$, or simply the coverage probability at $\bar{x}$. Thus, we can say that the asymptotic correctness of UB-M guarantees the asymptotic correctness at any optimal solution. Moreover, since $\vartheta^* \leq \mu(x)$ for all $x \in \bar{\mathcal{S}}_N^{\alpha}$, we obtain that

$$\Pr\left(\bar{\vartheta}_N^{\alpha} \geq \vartheta^* | \ \mathbb{P}\right) \geq (1-\alpha) + O\left(N^{-\kappa/2}\right).$$

This implies that the nominal confidence level approximately represents the lower bound of the probability that $\bar{\vartheta}_N^{\alpha}$ serves as an upper bound for ϑ^*. The following theorem shows the asymptotic correctness of APUB-M.

Theorem 3.14. Suppose that Assumptions 3.10 and 3.11 hold. Assume that the skewness of $F(x,\xi)$ is finite for each $x \in \mathcal{K}$. Then, APUB-M is 1 st-order asymptotically correct for $\alpha \in (0,1]$, i.e.,

$$\beta\left(\hat{\vartheta}_N^{\alpha}, \hat{\mathcal{S}}_N^{\alpha}\right) \geq (1-\alpha) + O\left(N^{-1/2}\right),$$

Proof. We know that $\hat{\mathcal{S}}_N^{\alpha} \subseteq \mathcal{K}$ w.p. 1 under Assumption 3.10 and the objective function of APUB-M, $\mathbb{U}_{\text{APUB?}}^{\alpha}\left[\mu(x) | \hat{\mathbb{P}}_N\right]$, is continuous by Proposition 3.12. Hence, $\hat{\mathcal{S}}_N^{\alpha}$ is compact for sufficiently large N w.p.1. Also, we know $\mu(x)$ is continuous. By the extreme value theorem, there exists $\tilde{x} \in \hat{\mathcal{S}}_N^{\alpha}$ such that

$$\mu(\tilde{x}) = \max_{x \in \mathcal{S}_N^{\alpha}} \mu(x).$$

Since the skewness of $F\left(\tilde{x}, \xi\right)$ is finite, it follows by Proposition 3.7 that

$$\begin{aligned}
\beta\left(\hat{\vartheta}_N^{\alpha}, \hat{\mathcal{S}}_N^{\alpha}\right) &= \beta\left(\mathbb{U}_{\text{APUB}}^{\alpha}\left[\mu\left(\tilde{x}\right) | \hat{\mathbb{P}}_N\right], \left\{\tilde{x}\right\}\right) \\
&= \Pr\left(\mathbb{U}_{\text{APUB}}^{\alpha}\left[\mu\left(\tilde{x}\right) | \hat{\mathbb{P}}_N\right] \geq \mu\left(\tilde{x}\right) | \mathbb{P}\right) \\
&\geq (1-\alpha) + O\left(N^{-1/2}\right).
\end{aligned}$$

Remark 3.15. The attribute of asymptotic correctness lends APUB-M interpretability in the context of statistics. This means that the decision-maker can intuitively set the desired reliability level of APUB-M by selecting an appropriate nominal confidence level. Section 3.5 provides a numerical demonstration of how this model interpretability confers an advantage.

3.3.2 Asymptotic Consistency

In optimization, asymptotic consistency refers to the convergence of the optimal value and optimal solution set of APUB-M with their counterparts in EM-M w.p. 1 as the sample size increases. The following theorem exhibits the asymptotic consistency of APUB-M.

Theorem 3.16. Suppose Assumptions 3.10 and 3.11 hold. Then, for any given $\alpha \in (0,1]$, as $N \to \infty$,

$$\hat{\vartheta}_N^{\alpha} \to \vartheta^*, \text{ and } \left(\hat{\mathcal{S}}_N^{\alpha}, \mathcal{S}\right) := \sup_{y \in \mathcal{S}_N} \inf_{z \in \mathcal{S}} \| y - z \| \to 0 \text{ w.p.1.}$$

Remark 3.17. Unlike DRO approaches that require additional parameter adjustments based on the sample size to achieve data-driven objectives, the sample size itself is the unique factor to determine the convergence of APUB-M. This characteristic offers a more consistent data-driven approach in practice. As ambiguity in distribution decreases with larger sample sizes, the influence of APUB consequently lessens. As a result, APUB-M avoids over-conservatism.

To prove Theorem 3.16, we first prove the following lemma, which shows the uniform consistency of $\mathbb{U}_{\text{APUB}}^{\alpha}\left[\mu(x) | \hat{\mathbb{P}}_N\right]$ on $\mathcal{K}$.

Lemma 3.18. Suppose Assumption 3.10 and 3.11 holds. Then, we have as $N \to \infty$,

$$\sup_{x \in \mathcal{K}} \left| \mathbb{U}_{\text{APUB}}^{\alpha}\left[\mu(x) | \hat{\mathbb{P}}_N\right] - \mu(x) \right| \to 0, \text{w.p.1.} \tag{3.11}$$

Proof. Without loss of generality, let the open convex set $\mathcal{N} \subseteq \mathbb{R}^d$. We first construct a countable dense subset of $\mathcal{N}$ as $\mathcal{D} := \mathbb{Q}^d \cap \mathcal{N}$, where $\mathbb{Q}^d$ represents the set of d dimensional rational numbers. Choose a sample path $(\xi_1, \xi_2, \ldots)$ and hence $\hat{\mathbb{P}}_N$ is the empirical distribution associated to the first N sample points. For $x \in \mathcal{D}$, we denote an event as

$$\mathrm{Y}_x := \left\{ (\xi_1, \xi_2, \ldots) : \lim_{N \to \infty} \mathbb{U}_{\text{APUB}}^{\alpha}\left[\mu(x) | \hat{\mathbb{P}}_N\right] = \mu(x) \right\}$$

Since $\mu(x) < \infty$ and $\sigma(x) < \infty$ under Assumption (B2), it follows by Theorem 3.8 that $\Pr\left(\mathrm{Y}_x | \mathbb{P}\right) = 1$, which implies that $\Pr\left(\cap_{x \in \mathcal{D}} \mathrm{Y}_x | \mathbb{P}\right) = 1$. In other words, $\mathbb{U}_{\text{APUB?}}^{\alpha}\left[\mu(x) | \hat{\mathbb{P}}_N\right]$ converges pointwisely to $\mu(x)$ on $\mathcal{D}$ w.p.1. Furthermore, by Proposition 3.12 and Theorem A. 7

[Rockafellar, 2015], we can conclude that $\mathbb{U}^{\alpha}_{\text{APUB?}}\left[\mu(x)|\ \hat{\mathbb{P}}_N\right]$ converges uniformly a certain continuous function ν on $\mathcal{K}$ w.p.1. Since $\nu(x)$ and $\mu(x)$ coincidence on a dense subset of $\mathcal{K}$ and they are both continuous on $\mathcal{K}$, we know that $\nu(x)=\mu(x)$ for all $x\in\mathcal{K}$. This completes the proof.

Proof of Theorem 3.18.

i) Proof of the consistency of $\hat{\vartheta}^{\alpha}_N$. Choose $x^*\in\mathcal{S}$ and $\hat{x}_N\in\hat{\mathcal{S}}_N$. It is easy to see that

$$\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(\hat{x}_N)|\ \hat{\mathbb{P}}_N\right]\le\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(x^*)|\ \hat{\mathbb{P}}_N\right]$$

and

$$\mu(x^*)\le\mu(\hat{x}_N)$$

Thus, we have

$$\begin{aligned}\left|\hat{\vartheta}^{\alpha}_N-\vartheta^*\right| &= \left|\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(\hat{x}_N)|\ \hat{\mathbb{P}}_N\right]-\mu(x^*)\right| \\ &= \max\left\{\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(\hat{x}_N)|\ \hat{\mathbb{P}}_N\right]-\mu(x^*),\mu(x^*)-\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(\hat{x}_N)|\ \hat{\mathbb{P}}_N\right]\right\} \\ &\le \max\left\{\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(x^*)|\ \hat{\mathbb{P}}_N\right]-\mu(x^*),\mu(\hat{x}_N)-\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(\hat{x}_N)|\ \hat{\mathbb{P}}_N\right]\right\} \\ &\le \sup_{x\in\mathcal{K}}\left|\mathbb{U}^{\alpha}_{\text{APUB}}\left[\mu(x)|\ \hat{\mathbb{P}}_N\right]-\mu(x)\right|\end{aligned}$$

which converges to 0 w.p. 1 by Theorem 3.18. This completes the proof.

ii) Proof of the consistency of $\hat{\mathcal{S}}^{\alpha}_N$. Let $\mathcal{O}$ as a collection of sample paths along which $\hat{\mathcal{S}}^{\alpha}_N\subseteq\mathcal{K}$ for a sufficiently large N and $\hat{\vartheta}^{\alpha}_N\to\vartheta^*$. By the above proof and Assumption (A2), we have $\Pr(\mathcal{O}|\ \mathbb{P})=1$. We now choose $(\xi_1,\xi_2,\ldots)\in\mathcal{O}$. Thus $\hat{\mathcal{S}}^{\alpha}_N$ is the optimal solution set of APUB-M using the first N sample points.

Suppose by contradiction that $\mathbb{D}\left(\hat{\mathcal{S}}_N^\alpha, \mathcal{S}\right) \nrightarrow 0$ along the sample path $(\xi_1, \xi_2, \ldots)$. Then, there exists $\varepsilon > 0$ such that for all $M \in \mathbb{N}$, there exists some $N > M$ for which $\mathbb{D}\left(\hat{\mathcal{S}}_N^\alpha, \mathcal{S}\right) > \varepsilon$. Specifically, there exists $\hat{x}_N \in \hat{\mathcal{S}}_N^\alpha$ such that $\inf_{y \in \mathcal{S}} \hat{x}_N, y > \varepsilon$. Because of the compactness of $\mathcal{K}$, we can find a subsequence $\hat{x}_{N_k} \in \hat{\mathcal{S}}_{N_k}^\alpha$ such that $\hat{x}_{N_k} \subseteq \mathcal{K}$ for all $k \in \mathbb{N}$, and

$$\lim_{k \to \infty} \hat{x}_{N_k} = \hat{x} \in \mathcal{K}, \inf_{y \in \mathcal{S}} \hat{x}_N, y > \varepsilon, \text{for all } k.$$

It follows that $\hat{x} \notin \mathcal{S}$ and hence $\mu(\hat{x}) > \vartheta^*$. On the other hand, we have

$$\left| \mathbb{U}_{\text{APUB}}^\alpha \left[\mu\left(\hat{x}_{N_k}\right) | \ \hat{\mathbb{P}}_N \right] - \mu(\hat{x}) \right| \leq \left| \mathbb{U}_{\text{APUB}}^\alpha \left[\mu\left(\hat{x}_{N_k}\right) | \ \hat{\mathbb{P}}_N \right] - \mu\left(\hat{x}_{N_k}\right) \right| + \left| \mu\left(\hat{x}_{N_k}\right) - \mu(\hat{x}) \right|.$$

On the right hand of the above inequality, the first term converges to zero by Theorem 3.18, and the second term converges to zero because of the continuity of $\mu(x)$. Thus,

$$\lim_{k \to \infty} \mathbb{U}_{\text{APUB}}^\alpha \left[\mu\left(\hat{x}_{N_k}\right) | \ \hat{\mathbb{P}}_N \right] = \mu(\hat{x}).$$

The definition of $\mathcal{O}$ ensures that $\mathbb{U}_{\text{APUB}}^\alpha \left[\mu\left(\hat{x}_{N_k}\right) | \ \hat{\mathbb{P}}_N \right] = \hat{\vartheta}_N^\alpha \to \vartheta^*$. It implies that $\mu(\hat{x}) = \vartheta^*$. This is contradictory to the assertion that $\mathbb{D}\left(\hat{\mathcal{S}}_N^\alpha, \mathcal{S}\right) \nrightarrow 0$.

3.4 Solution Method Based on Sampling Approximation

By Proposition 3.4, we rewrite APUB-M as

$$\hat{\vartheta}_N^\alpha = \min_{(x,t) \in \mathcal{X} \times \mathbb{R}} t + \frac{1}{\alpha} \int \left[\frac{1}{N} \sum_{n=1}^N F(x, \zeta_n) - t \right]^+ \prod_{n=1}^N \hat{\mathbb{P}}_N(d\zeta_n). \tag{3.12}$$

Let $\hat{\mathcal{Z}}_N^{\alpha}$ be the optimal solution set of (3.12). We enumerate the all permutations of the sample points $(\xi_1,\ldots\xi_N)$ with replacement, denoted by $(\zeta_{m,1},\ldots,\zeta_{m,N})$, for $m=1,\ldots,M$ where $M=N^N$. APUB-M can be represented as

$$\min_{(x,t)\in\mathcal{X}\times\mathbb{R}} t+\frac{1}{\alpha M}\sum_{m=1}^{M}\left[\frac{1}{N}\sum_{n=1}^{N}F(x,\zeta_{m,n})-t\right]_+. \qquad \text{(BP-APUB-M)}$$

The above model representation comprises N^{N+1} random scenarios. Addressing its large-scale nature becomes imperative. To tackle the complexity arising from the sheer number of scenarios, we can leverage the sampling approximation method. Achieving a satisfactory approximation, BP-APUB-M actually needs a significantly smaller number $M \ll N^N$ of random samples. Each sample, drawn from the empirical distribution $\hat{\mathbb{P}}_N$, consists of N sample points. The random samples are referred to as bootsrap samples, in terms of the nonparametric bootstrap percentile method.

3.4.1 Asymptotic Convergence of BP-APUB-M

This section is in reference to the asymptotic behavior of BP-APUB-M as the N original sample points $(\xi_1,\ldots,\xi_N)$ are fixed and the number M of bootstrap samples increases. Theorem 3.19 shows the convergence of BP-APUB-M, and Theorem 3.21 explores its stability.

Theorem 3.19. Suppose that $\mathcal{X}$ is compact and $F(x,\xi_n)$ is continuous convex on $\mathcal{X}$ for any orginal sample point $\xi_n, n=1,\ldots,N$. Let $\tilde{\vartheta}_{N,M}^{\alpha}$ and $\tilde{\mathcal{Z}}_{M,N}^{\alpha}$ denote the optimal value and the solution set of $BP-APUB-M$, respectively. Then, for any $\alpha\in(0,1]$, as $M\to\infty$,

$$\tilde{\vartheta}_{M,N}^{\alpha}\to\hat{\vartheta}_N^{\alpha} \text{ and } \mathbb{D}\left(\tilde{\mathcal{Z}}_{M,N}^{\alpha},\hat{\mathcal{Z}}_N^{\alpha}\right)\to 0 \text{w.p.1}(\text{for}\zeta).$$

To prove Theorem 3.19, we first give the following Lemma.

Lemma 3.20. Let the assumptions of Theorem 3.19 hold. Denote

$$t_u := \sup_{x\in\mathcal{X},\zeta\in\{\xi_1,\ldots,\xi_N\}} F(x,\zeta) \text{ and } t_l := \inf_{x\in\mathcal{X},\zeta\in\{\xi_1,\ldots,\xi_N\}} F(x,\zeta)$$

Then, t_u and t_l are finite. Let $\mathcal{T} := [t_l, t_u], B := \frac{2+\alpha}{\alpha}\max\{|t_u|,|t_l|\}$, and $(\eta_1,\ldots,\eta_N) \sim \hat{\mathbb{P}}_N$ be a generic bootstrap sample. Then,

$$\sup_{(x,t)\in\mathcal{X}\times\mathcal{T}} \left| t + \frac{1}{\alpha}\left[\frac{1}{N}\sum_{n=1}^{N} F(x,\eta_n) - t\right]_+ \right| \le B \text{w.p.1}(\text{for}\,\eta).$$

Furthermore, for any $\alpha \in (0,1]$, $BP-APUB-M$ is equivalent to

$$\min_{(x,t)\in\mathcal{X}\times\mathcal{T}} t + \frac{1}{\alpha M}\sum_{m=1}^{M}\left[\frac{1}{N}\sum_{n=1}^{N} F(x,\zeta_{m,n}) - t\right]_+ \text{ w.p.1}(\text{for}\zeta), \tag{3.13}$$

which substitutes $\mathcal{T}$ for $\mathbb{R}$ *in BP-APUB-M.*

Proof. Since $F(x,\xi_n)$ is continuous on the compact set $\mathcal{X}$ for all $\xi_n, n = 1,\ldots,N$. It follows by the extreme value theorem that t_l and t_u are bounded. Then, we have that, for all $(x,t)\in\mathcal{X}\times\mathcal{T}$,

$$\left| t + \frac{1}{\alpha}\left[\frac{1}{N}\sum_{n=1}^{N} F(x,\eta_n) - t\right]_+ \right| \le |t| + \frac{1}{\alpha}\left|\frac{1}{N}\sum_{n=1}^{N} F(x,\eta_n) - t\right| \le \frac{2+\alpha}{\alpha}\max\{|t_u|,|t_l|\} = B \text{ w.p.1.}$$

To prove the equivalence between BP-APUB-M and model (3.13), we only need to show that the optimal solution set of BP-APUB-M is contained in $\mathcal{X}\times\mathcal{T}$. The key of the proof is to show that any optimal t-solution belongs to $\mathcal{T}$. It is clear that, for an arbitrary $\delta > 0$ and all $x\in\mathcal{X}$,

$$\frac{1}{N}\sum_{n=1}^{N} F(x,\eta_n) \ge t_l > t_l - \delta := t' \text{w.p.1.}$$

It follows that, for any $x\in\mathcal{X}$,

$$\left\{t_l+\frac{1}{\alpha M}\sum_{m=1}^{M}\left[\frac{1}{N}\sum_{n=1}^{N}F\left(x,\zeta_{m,n}\right)-t_l\right]_+\right\}-\left\{t'+\frac{1}{\alpha M}\sum_{m=1}^{M}\left[\frac{1}{N}\sum_{n=1}^{N}F\left(x,\zeta_{m,n}\right)-t'\right]_+\right\}=\delta-\frac{\delta}{\alpha}<0,\text{w.p.1.}$$

Since δ is arbitrary, any optimal t-solution of BP-APUB-M should not be less than t_l.

On the other hand, a similar proof can show that any optimal t-solution of BP-APUBM should not be more than t_u.

Proof of Theorem 3.19. Under the assumption, it is easy to see that

$$t+\frac{1}{\alpha}\left[\frac{1}{N}\sum_{n=1}^{N}F\left(x,\eta_n\right)-t\right]_+$$

is continuous on $\mathcal{X}\times\mathbb{R}$ w.p.1. By Lemma 3.20, we also know that for any $\left(x,t\right)\in\mathcal{X}\times\mathcal{T}$,

$$\left|t+\frac{1}{\alpha}\left[\frac{1}{N}\sum_{n=1}^{N}F\left(x,\eta_n\right)-t\right]_+\right|\le B<\infty\text{w.p.1.}$$

BP-APUB-M satisfies the conditions required in Theorems A. 9 [Shapiro et al., 2021]. Subsequently, we have that as $M\to\infty$,

$$\sup_{(x,t)\in\mathcal{X}\times\mathbb{R}}\left|\frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{N}\sum_{n=1}^{N}F\left(x,\zeta_{m,n}\right)-t\right]_+-\int\left[\frac{1}{N}\sum_{n=1}^{N}F\left(x,\eta_n\right)t\prod_{+}^{N}\hat{\mathbb{P}}_N\left(d\eta_n\right)\right|\to 0,\text{w.p.1.}$$

Then, by Theorem A. 10 [Shapiro et al., 2021], we can complete the proof.

Theorem 3.21. Let the assumptions of Theorem 3.19 hold. Let $\left(\tilde{x}_{M,N}^{\alpha},t_{M,N}^{\alpha}\right)$ *be an optimal solution of* $BP-APUB-M$. Then, for any $\varepsilon>0$, there exist $\mathfrak{a}_\varepsilon>0$ and $\mathfrak{b}_\varepsilon>0$ *such that*

$$\Pr\left(\mathbb{U}_{APUB}^{\alpha}\left[\mu\left(\tilde{x}_{M,N}^{\alpha}\right)|\ \hat{\mathbb{P}}_N\right]-\hat{\theta}_N^{\alpha}>\varepsilon|\ \Pi^N\left(\hat{\mathbb{P}}_N\right)\right)\le\mathfrak{a}_\varepsilon e^{-\mathfrak{b}_\varepsilon M},$$

for all $M>1$. If $\hat{\mathcal{Z}}_N^{\alpha}=\left\{\left(\hat{x}_N^{\alpha},\hat{t}_N^{\alpha}\right)\right\}$ is a singleton, there exist $\mathfrak{a}_\varepsilon'>0$ and $\mathfrak{b}_\varepsilon'>0$ such that

$$\Pr\left(\left(\tilde{x}_{M,N}^{\alpha},\tilde{t}_{M,N}^{\alpha}\right)-\left(\hat{x}_N^{\alpha},\hat{t}_N^{\alpha}\right)>\epsilon|\ \Pi^N\left(\hat{\mathbb{P}}_N\right)\right)\le\mathfrak{a}_\varepsilon' e^{-\mathfrak{b}_\varepsilon' M},$$

for all $M > 1$.

Proof. The proof of this theorem directly follows from Lemma 3.20 and Theorem A. 11 [Birge and Louveaux, 2011].

3.4.2 Practical Reformulation of BP-APUB-M

Recall that $(\xi_1, \ldots, \xi_N)$ is the original random sample associated with the empirical distribution $\hat{\mathbb{P}}_N$. In the context of nonparametric bootstrap sampling, each point in a bootstrap sample is drawn from $(\xi_1, \ldots, \xi_N)$ with replacement. We can count the number of times the specific original sample point ξ_n appears in the m th bootstrap sample $(\zeta_{m,1}, \ldots, \zeta_{m,N})$, which is denoted by $V_{m,n}$. Note that $0 \le V_{m,n} \le N$. This observation implies that BP-APUB-M can be reformulated as

$$\min_{(x,t)\in\mathcal{X}\times\mathbb{R}} t + \frac{1}{\alpha M}\sum_{m=1}^{M}\left[\frac{1}{N}\sum_{n=1}^{N} V_{m,n} F(x, \xi_n) - t\right]_+ . \tag{3.14}$$

As an illustrative application, we now demonstrate the utility of APUB within the context of a specific class of optimization problems known as two-stage linear stochastic optimization with random recourse. In this case, APUB-M is adapted to formulate the first stage as

$$\min_x c^\top x + \mathbb{U}^{\alpha}_{\text{APUB}}\left[\mathbb{E}_{\mathbb{P}}\left[Q(x,\xi)\right] \middle| \hat{\mathbb{P}}_N\right] \tag{3.15a}$$

$$\text{s.t.} Ax = b, \tag{3.15b}$$

$$x \ge 0. \tag{3.15c}$$

By letting $\xi = (q, h, T, W)$, the second stage is represented as

$$Q(x,\xi) = \min_y q^\top y \tag{3.16a}$$

$$\text{s.t.} Wy = h - Tx, \tag{3.16b}$$

$$y \geq 0. \tag{3.16c}$$

Denote $\eta(x) := c^\top x + \mathbb{U}^{\alpha}_{\text{APUB?}}\left[\mathbb{E}_{\mathbb{P}}\left[Q(x,\xi)\right]|\ \hat{\mathbb{P}}_N\right]$ and $\mathfrak{X} := \{x : \eta(x) < \infty, Ax = b, x \geq 0\}$.

Furthermore, we equivalently write the first stage (3.15) as

$$\min_{x \in \mathcal{X}} c^\top x + \mathbb{U}^{\alpha}_{\text{APUB}}\left[\mathbb{E}_{\mathbb{P}}\left[Q(x,\xi)\right]|\ \hat{\mathbb{P}}_N\right], \tag{3.17}$$

where the relatively complete recourse is satisfied, i.e., $Q(x,\xi_n) < \infty$ for all $x \in \mathfrak{X}$ and $n = 1,\ldots,N$. Relatively complete recourse is commonly a reasonable condition, especially given that a solution lacking feasible recourse action can generally be deemed ill-defined [?].

Let $\xi_n = (q_n, h_n, T_n, W_n)$ be the n th point of the original random sample associated with the empirical distribution $\hat{\mathbb{P}}_N$. The bootstrap sampling approximation of the two-stage APUB-M (3.16)-(3.17) can be written as a linear program,

$$\min_{x,y,s,t} c^\top x + t + \frac{1}{\alpha M}\sum_{m=1}^{M} s_m \tag{3.18a}$$

$$\text{s.t.} s_m \geq -t + \frac{1}{N}\sum_{n=1}^{N} V_{m,n} q_n^\top y_n, \qquad m = 1,\ldots,M, \tag{3.18b}$$

$$W_n y_n = h_n - T_n x, \qquad n = 1,\ldots,N, \tag{3.18c}$$

$$Ax = b, \tag{3.18d}$$

$$x \geq 0, y \geq 0, s \geq 0 \tag{3.18e}$$

In Sections 3.5.1 and 3.5.2, we test the performance of the two-stage APUB-M, comparing with SAA-M and DRO approaches.

3.5 Numerical Analyses

We assess the efficacy of APUB-M through an extensive examination of classic problems in stochastic optimization, spanning both single-stage and two-stage scenarios. Section 3.5.1 provides a comparative analysis between APUB-M and traditional DRO utilizing Wasserstein distance. This comparison involves evaluating their respective out-of-sample performances and coverage probabilities in addressing a two-stage benchmark problem with fixed recourse (Dantzig, 2016). The comparative analysis reveals that, although APUB-M does not guarantee 100% coverage probability in situations characterized by a severe lack of data, it demonstrates a potential advantage by mitigating over-conservatism and achieving better average out-of-sample performance than the DRO approach. Furthermore, Section 3.5.2 extends the application scope of APUB-M to encompass problems featuring random recourse. It shows that the robustness and favorable performance of APUB-M are maintained. Traditional DRO methodologies encounter inherent limitations when confronted with the computational complexity resulting from random recourse. In Section 3.5.3, we subject APUB-M to rigorous testing using a multi-product newsvendor problem (Hanasusanto et al., 2015). APUB-M provides stable and high-quality solutions even when the sample size is small. Notably, Mohajerin Esfahani and Kuhn (2018) highlight that the Wasserstein distance based DRO fails to perform effectively in situations, like the newsvendor problem, where the random loss function exhibits a Lipschitz modulus concerning random scenarios, independent of decision variables. Across all investigated scenarios, the close correspondence between nominal confidence levels and actual coverage probabilities serves as a testament to the methodological reliability of APUB-M.

3.5.1 A Two-Stage Product Mix Problem with Fixed Recourse

We adapt the benchmark two-stage product mix problem presented by Dantzig [2016] to our test case, which seeks to optimize the product mix of a furniture shop amid uncertain labor conditions. During the 'here-and-now' stage, the company commits to a long-term contract, promising to deliver a set quantity of furniture in each time period. This quantity can be adjusted due to strong market demand. Labor hours, which are crucial to production, are uncertain and variable, partly because of factors such as the COVID-19 pandemic. The production involves four distinct products and two workstations each constrained by the availability of labor hours. Each product requires varying amounts of labor across these workstations and contributes specific profit margins upon sale. Labor availability dictates production time, with more hours leading to reduced production time, a phenomenon attributed to skill diversity and improved efficiency. At this stage, the company's objective is to determine the most profitable product mix that meets contractual requirements while contending with the unpredictability of labor availability.

In the subsequent 'wait-and-see' stage, the company must confront the actual labor hours available, which may deviate from earlier estimates. In instances where there is a shortfall in the labor hours anticipated by the production plan conceived in the "here-and-now' stage, the option exists to outsource additional labor hours for workstation. However, this supplemental workforce is not as efficient as the in-house labor. Thus, at this juncture, the firm's focus pivots to minimizing the expenses linked to acquiring these supplemental, less efficient labor hours, while still fulfilling the contractual furniture delivery commitments. The decision-making process in this stage is heavily dependent on actual labor availability and is geared toward economical adjustments to labor shortages.

In practice, the company determines the production quantities outlined in the contract by analyzing historical labor hours. However, the unpredictability of absenteeism, exacerbated by the

COVID-19 pandemic, has led to a significant lack of reliable data. In response to this uncertainty, the company seeks to define a profit threshold that the expected profit from this contract is likely to meet or exceed, maintaining a confidence level of approximately $100(1-\alpha)\%$. We describe the profit threshold in form of APUB, which ensures the statistical reliability of the company's objective. On this basis, we represent this product mix problem as the two-stage APUB-M (3.15)-(3.16). In the first stage, x is the decision vector for the product mix and the negative value of c represents the per-unit profits of products. In the second stage, the decision vector y signifies the outsourced labor hours assigned to workstations and associated with unit cost q (the last two components with a cost of zero correspond to two slack variables), h stands for the random labor hours available at workstations, T includes the production times required for products, and the negative value of W represents the efficiency rate of outsourced labor. The numerical parameters in our study are specified as follows:

$$
\begin{aligned}
A &= 0, b = 0, c = [-12, -20, -18, -40]^\top \\
q &= [6, 12, 0, 0]^\top, h = \left[500\gamma_1, 500\gamma_2\right]^\top, \\
T &= \begin{bmatrix} 4-\frac{\gamma_1}{4} & 9-\frac{\gamma_1}{4} & 7-\frac{\gamma_1}{4} & 10-\frac{\gamma_1}{4} \\ 3-\frac{\gamma_2}{4} & 1-\frac{\gamma_2}{4} & 3-\frac{\gamma_2}{4} & 6-\frac{\gamma_2}{4} \end{bmatrix}, W = \begin{bmatrix} -0.9 & 0 & 1 & 0 \\ 0 & -0.9 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

where

$$
\left[\gamma_1, \gamma_2\right]^\top \sim \frac{7}{10}\mathcal{N}\left(\begin{bmatrix} 12 \\ 8 \end{bmatrix}, \begin{bmatrix} 5.76 & 1.92 \\ 1.92 & 2.56 \end{bmatrix}\right) + \frac{3}{10}\mathcal{N}\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0.16 & 0.04 \\ 0.04 & 0.04 \end{bmatrix}\right)
$$

has a mixed 2-dimensional normal distribution. In the two-stage APUB-M (3.15)(3.16), we generate a random sample with a size N from the mixed normal distribution.

3.5.1.1 Convergence of Bootstrap Sampling Approximation.

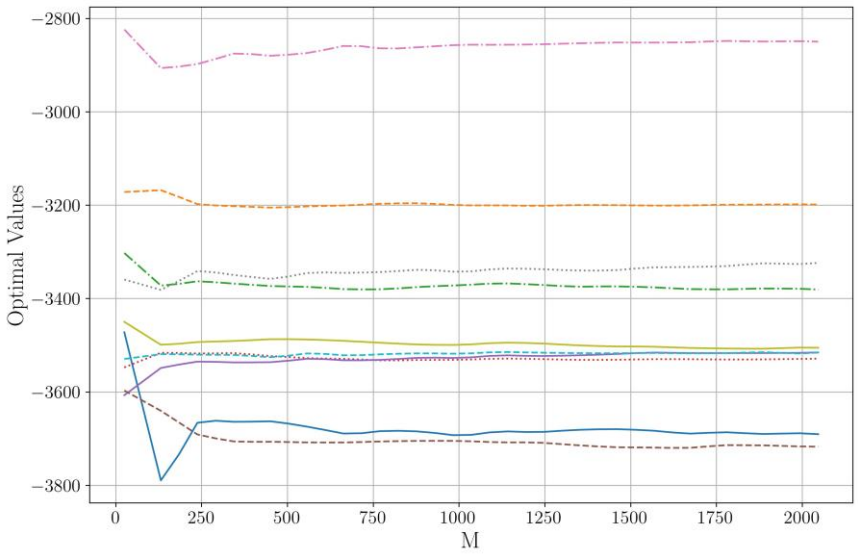
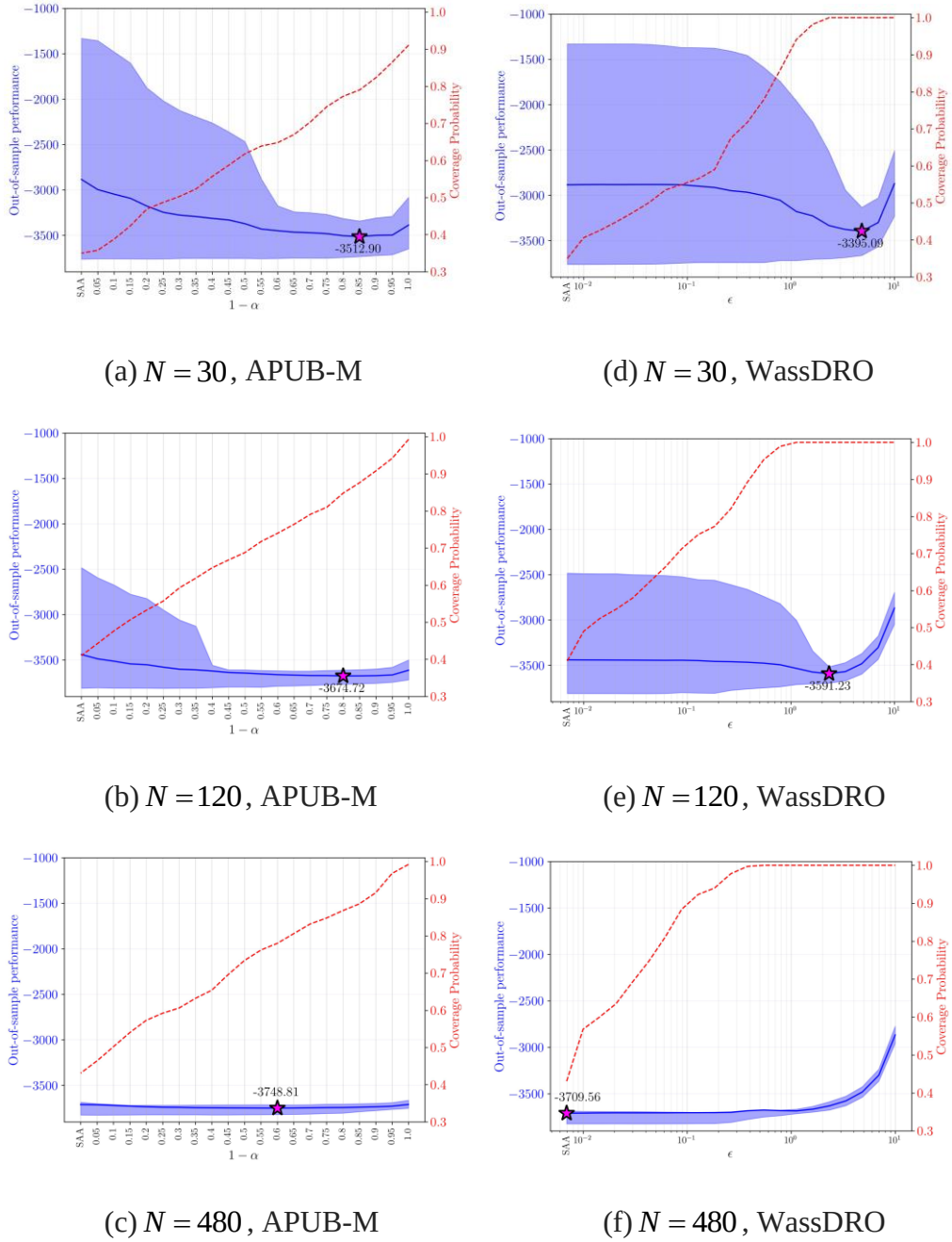


Figure 3.3: Convergence of the bootstrap sampling approximation.

We now assess the convergence of the bootstrap sampling method applied to our model. Our evaluation encompasses 10 independent simulations, each producing $N = 30$ sample data points and resolving the subsequent approximation as defined in (3.18). Throughout these tests, we maintain a consistent nominal confidence level of $(1-\alpha) = 0.8$. Figure 3.3 illustrates the relationship between the number M of bootstrap samples and the optimal values of our approximation problem, with M reaching up to 2000. A discernible stabilization trend is evident in the data: as M increases, variability in the optimal values conspicuously decreases. Notably, for $M \geq 1000$, the convergence of the approximation becomes evident as the fluctuation in the optimal values significantly lessens. This consistency bolsters our decision to adopt $M = 2000$ for all subsequent experiments in this section.



(a) $N = 30$, APUB-M (d) $N = 30$, WassDRO

(b) $N = 120$, APUB-M (e) $N = 120$, WassDRO

(c) $N = 480$, APUB-M (f) $N = 480$, WassDRO

Figure 3.4: Out-of-sample performance (left axis, solid line, and shaded area) and the coverage probability (right axis, dashed line) as a function of the nominal confidence level $(1-\alpha)$ in APUB-M and a function of ϵ in WassDRO. The star symbol indicates the point where the mean of the out-of-sample performance attains its minimum. The minimum value of the mean is written next to the star symbol.

3.5.1.2 Comparative Analysis between APUB-M and DRO.

We carry out a comparative analysis between APUB-M and the Wasserstein distance based DRO approach when applied to this two-stage product mix problem. Recall that, used in the two-stage problem, APUB-M is formed as (3.15)-(3.16). Our aim is to evaluate their performance under various conditions and identify their respective strengths and limitations.

All the tests are conducted using a Monte Carlo simulation of 1000 replications. In each replication, we generate a training sample set of size N, using which APUB-M provides an optimal solution $\hat{x}_N^\alpha$. The out-of-sample performance, $\mu\left(\hat{x}_N^\alpha\right)$, is evaluated using a unique large-sized test sample set. The all replications provide the approximate probability distribution of $\mu\left(\hat{x}_N^\alpha\right)$. For $N = 30, 120$, and 480, respectively, Figures 3.4a3.4c display the curve of the mean of $\mu\left(\hat{x}_N^\alpha\right)$ and the range from the 10 th to the 90 th percentile when varying $(1-\alpha)$ from 0 to 1. Note that the leftmost case represents the out-of-sample performance of SAA-M, which is equivalently represented as APUB-M with $(1-\alpha) = 0$. We also estimate the coverage probability of APUB-M concerning $\hat{x}_N^\alpha$ as

$$\beta\left(\hat{\vartheta}_N^\alpha, \left\{\hat{x}_N^\alpha\right\}\right) = \Pr\left(\hat{\vartheta}_N^\alpha \geq \mu\left(\hat{x}_N^\alpha\right) | \ \mathbb{P}\right).$$

Recall that the function β is defined in (3.10). By Theorem 3.14, we know the asymptotic correctness of APUB-M concerning $\hat{x}_N^\alpha$, i.e.,

$$\beta\left(\hat{\vartheta}_N^\alpha, \left\{\hat{x}_N^\alpha\right\}\right) \geq (1-\alpha) + O\left(N^{-1/2}\right).$$

Figures 3.4a-3.4c draw the curve of $\beta\left(\hat{\vartheta}_N^{\alpha}, \left\{\hat{x}_N^{\alpha}\right\}\right)$ with respect to $(1-\alpha)$. Similarly, Figures 3.4d – 3.4f report the out-of-sample performance and the coverage probability of the Wasserstein distance based DRO (labelled WassDRO in the figures), when altering the radius ϵ of the ball in 1-Wasserstein distance metric.

Comparing the out-of-sample performances and coverage probabilities of the two approaches in Figure 3.4, we have the following observations:

Analysis of the minimum average costs (where negative values represent profits), as indicated by stars, and their associated 90th percentiles across varying sample size N, consistently showed that APUB-M achieves lower costs in comparison to WassDRO. Despite this, WassDRO demonstrates higher coverage probabilities when achieving its minimum average costs. This observation indicates a potential strategic compromise between minimizing the cost objective and improving the reliability of the solution.

For small and medium sample sizes $(N = 30$ an $N = 120)$, both APUB-M and WassDRO substantially outpace SAA-M. This advantage is evident from the notably lower and more focused cost distributions over a broad spectrum of nominal confidence levels and radii. Additionally, both APUB-M and WassDRO reveal improvements in coverage probability, lending further support to their robustness and operational efficacy.

For medium and large sample sizes $(N = 120$ an $N = 480)$, we observe an accurate alignment of nominal confidence level and actual coverage probabilities attests to the methodological soundness of APUB-M. This alignment not only acts as a validation of the fidelity of the method in asymptotic correctness but also highlights a significant methodological stride in the interpretability of intuitively chosen reliability levels.

In the large sample scenarios $(N = 480)$, while an increase in ϵ within the WassDRO framework improves the coverage probability, the associated cost is concurrently magnified, especially when ϵ is selected to be excessively large. This phenomenon suggests that an inappropriately chosen ϵ can lead to significant over-conservatism. In contrast, APUB-M sustains the out-of-sample performance at a level analogous to SAA-M, while exhibiting a consistent increment in coverage probability. Such findings validate the stability of APUB-M, affirming that the asymptotic consistency of APUB-M is maintained irrespective of the chosen nominal confidence level.

Lastly, it is noteworthy that, for both APUB-M and WassDRO, the critical $(1-\alpha)$ and ϵ points, corresponding to the minimum average costs, approach the leftmost point representing SAA-M as N increases. This observation accentuates the benefits of SAA-M in a large sample scenario, where distributional ambiguity is reduced.

Overall, APUB-M not only exemplifies robustness but also avoids the over-conservatism often seen in traditional DRO approaches. Crucially, the nominal confidence level $(1-\alpha)$ functions as a faithful reflection of its statistical meaning, aligning with the company's preference for a specific confidence level amidst distributional ambiguity. This functionality bolsters the statistical interpretability of our method, providing clear, relevant insights even before model training commences. However, it is important to recognize that, in contrast to DRO, APUB-M does not invariably guarantee a 100% coverage probability. This limitation is particularly evident in cases with too few data points, where even the worst-case scenarios in the sample may fail to encompass the most extreme eventualities, an issue that becomes pronounced in the face of severe sample scarcity.

3.5.2 A Two-Stage Product Mix Problem with Random Recourse

We now extend our analysis to a two-stage stochastic optimization problem incorporating random recourse. To facilitate direct comparison, we modify the test problem outlined in Section 3.5.1 to incorporate this element of randomness in the recourse measures. Accordingly, we define the random course as

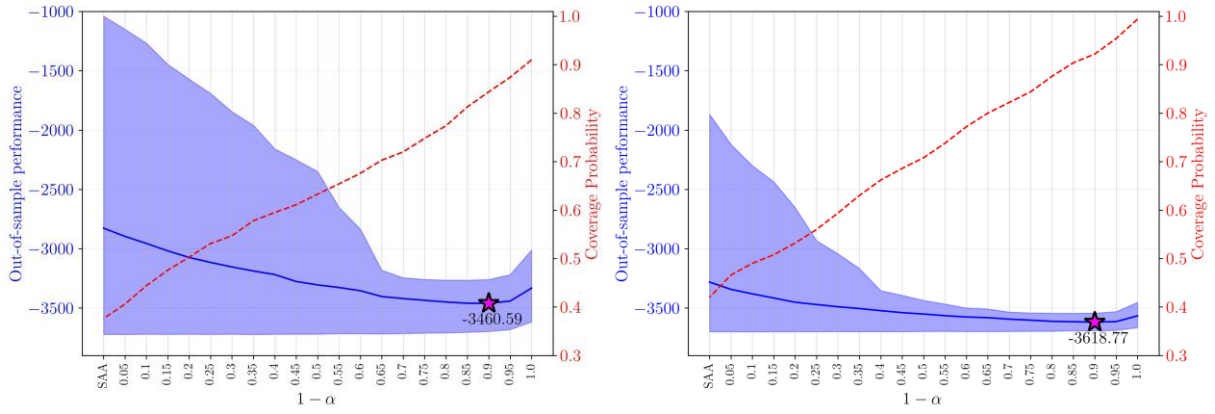
$$W = \begin{bmatrix} -\mathcal{U}(0.6,1.2) & 0 & 1 & 0 \\ 0 & -\mathcal{U}(0.6,1.2) & 0 & 1 \end{bmatrix}$$

where $\mathcal{U}$ represents a uniform distribution. In this product mix optimization, the random recourse refers to the treatment of externally sourced labor hours as stochastic variables in the second stage, reflecting the real-world variability in the labor market.

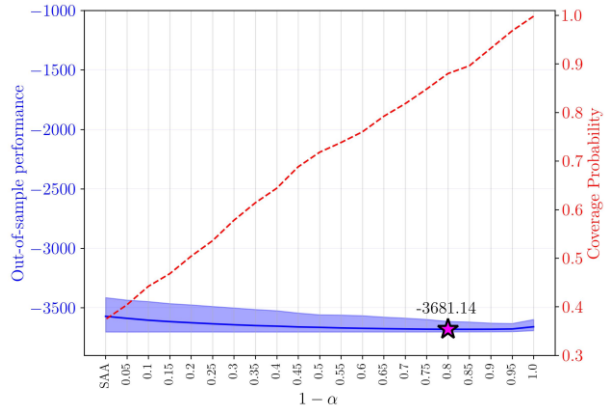
We first observe the out-of-sample performance of SAA-M in both fixed and random recourse scenarios. The results, depicted in Figures 3.5a through 3.5b, reveal that the random recourse scenario exhibits a wider 10-90th percentile range and more than a 300-unit increase in the 90th percentile for both $N = 30$ and $N = 120$, as well as a

higher mean, compared to the fixed recourse scenario (shown in Figures 3.4a through 3.4b). This indicates a heightened volatility in the random recourse case.

Nevertheless, APUB-M maintains a consistent performance profile in the random recourse scenario, as demonstrated in Figure 3.5, similar to its behavior in fixed recourse settings. This supports the model's methodological flexibility. Specifically, Figure 3.5a shows that APUB-M is exceptionally resilient when dealing with limited data, effectively reducing the mean of the cost and enhancing solution stability. This underscores the robustness of APUB-M when a suitable nominal confidence level is chosen. Additionally, Figure 3.5b confirms the asymptotic correctness of APUB-M. In



(a) $N = 30$ (b) $N = 120$



(c) $N = 480$

Figure 3.5: Out-of-sample performance (left axis, solid line, and shaded area) and the coverage probability (right axis, dashed line) as a function of the nominal confidence level $(1-\alpha)$ in APUB-M. The star symbol indicates the point where the mean of the out-of-sample performance attains its minimum. The minimum value of the mean is written next to the star symbol.

other words, APUB-M consistently meets the actual coverage probability when varying nominal confidence levels, paralleling its fixed recourse performance. Moreover, Figure 3.5c suggests that

with a large sample size $(N = 480)$, APUB-M can avoid excessive conservatism irrespective of the nominal confidence level.

On the other hand, in the random recourse scenario, using the same sample size, achieving the minimum average cost (signified by a star in Figure 3.5) requires a higher nominal confidence level to ensure greater robustness. Consequently, this minimum average cost is higher than that of the fixed recourse scenario. Additionally, the 10-90th percentile range is marginally broader across all nominal confidence levels, compared to the fixed recourse case. These findings indicate a necessity for setting a larger nominal confidence level in APUB-M to effectively manage the increased uncertainty introduced by random recourse.

3.5.3 A Multi-Product Newsvendor Problem

Consider a multi-product newsvendor problem, described in [Hanasusanto et al., 2015], with the following random cost function,

$$F(x,\xi) = p^{\top}x + h^{\top}(x-\xi)_{+} + b^{\top}(\xi - x)_{+},$$

where x is the vector of order quantities for ten products, ξ represents random demand, p is the unit profit ($p<0$ in the cost function), h and b are overage and underage costs.

3.5.3.1 Out-of-Sample Performance Analysis.

This test compares two cases: Case I assumes that ξ follows a mixed normal distribution as

$$\text{Case I:}\xi \sim \frac{1}{2}\mathcal{N}(\mu_1,\Sigma_1) + \frac{1}{2}\mathcal{N}(\mu_2,\Sigma_2),$$

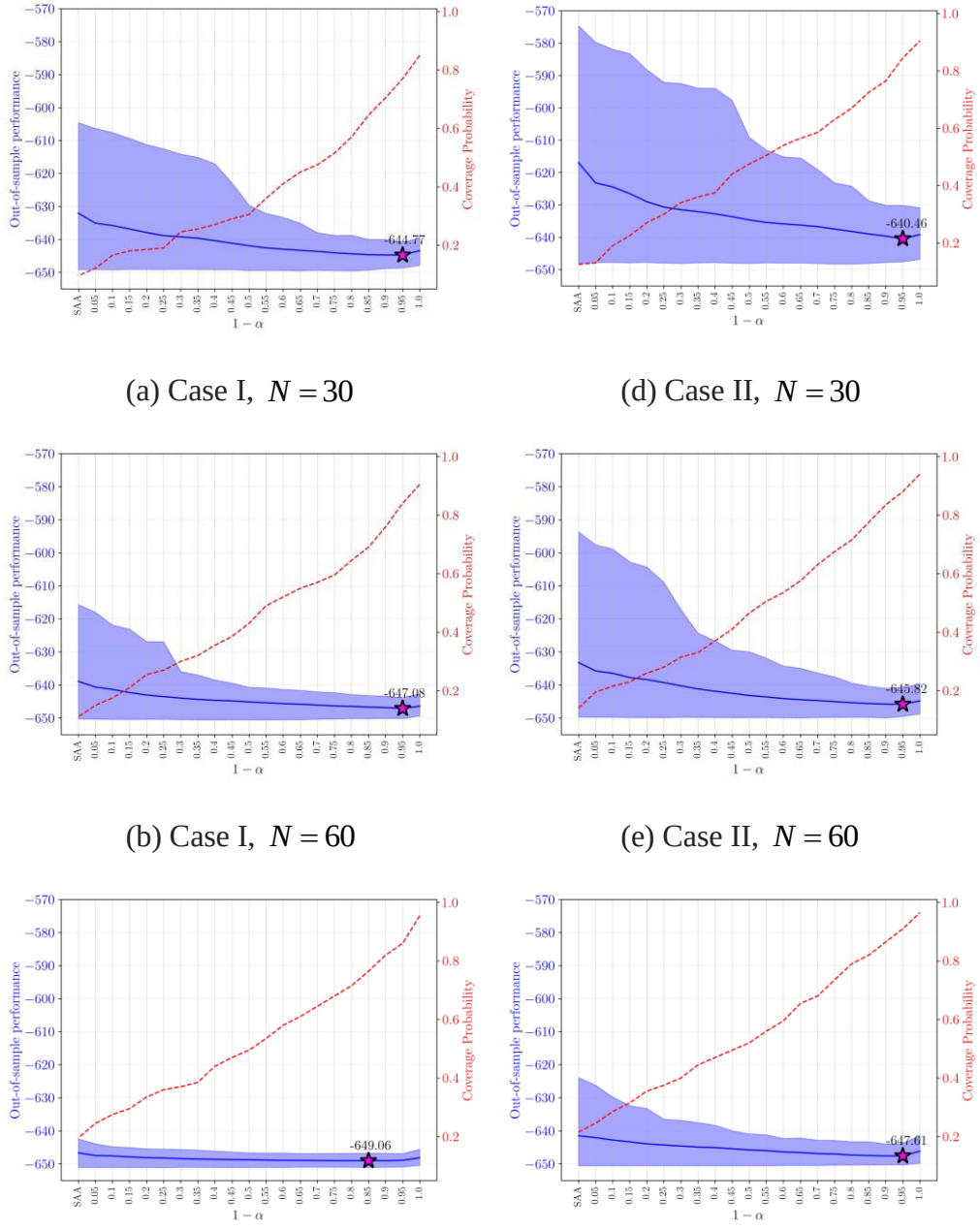
and on this basis, Case II considers a biased noise added in the data generation as

$$\text{Case II: } \xi' = \xi + \varepsilon.$$

With the noise ε included, Case II has a large variation and, as a result, its distributional ambiguity is more serious. The numerical values of all parameters are provided in Appendix B.

Figures 3.6a to 3.6c depict the out-of-sample performances and coverage probabilities of APUB-M in Case I as the sample size N varies from 30 to 120, while Figures 3.6d to 3.6f illustrate Case II. In both cases, APUB-M outperforms SAA-M in terms of lower average cost and a narrower range from the 10th to the 90th percentile. Additionally, Figures 3.6 showcase the asymptotic correctness and consistency of APUB-M, as observed in Sections 3.5.1 and 3.5.2.

On the other hand, SAA-M exhibits more stable performance in Case I than in Case II. This suggests that Case I experiences less distributional ambiguity. We observe that APUB-M is adaptable to different levels of ambiguity. However, the model's performance is sensitive to the level of ambiguity. When $N = 30$, Case II demonstrates a much wider 10-90th percentile range and requires a larger nominal confidence level $(1-\alpha)$ to achieve the minimum average cost indicated by a star. Increasing N weakens the impact of noise on distributional ambiguity. APUB-M with a higher $(1-\alpha)$ exhibits



(a) Case I, $N = 30$

(d) Case II, $N = 30$

(b) Case I, $N = 60$

(e) Case II, $N = 60$

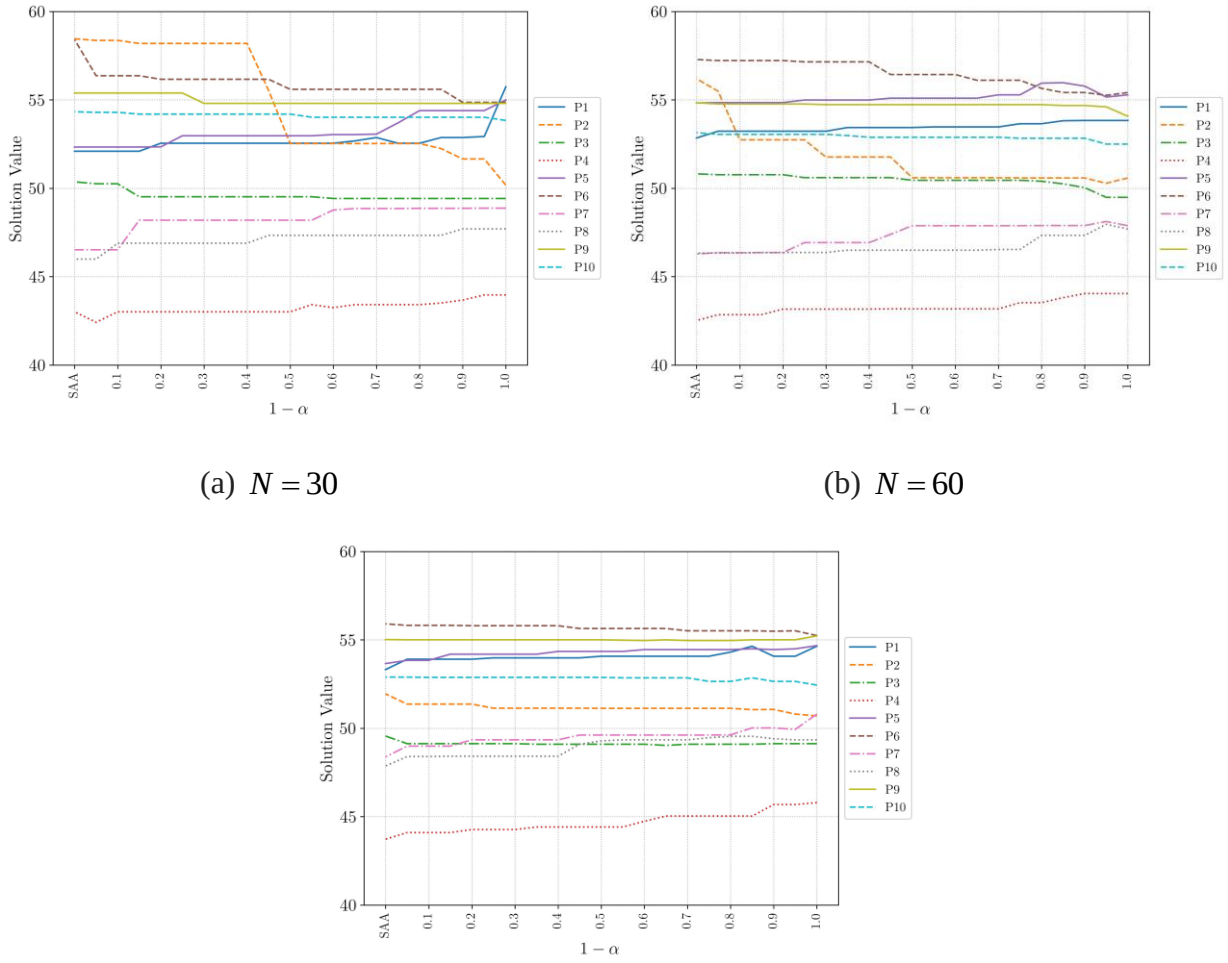
(c) Case I, $N = 120$

(f) Case II, $N = 120$

Figure 3.6: Out-of-sample performance (left axis, solid line, and shaded area) and the coverage probability (right axis, dashed line) as a function of the nominal confidence level $(1-\alpha)$ in APUB-M. The star symbol indicates the point where the mean of the out-of-sample performance attains its minimum. The minimum value of the mean is written next to the star symbol.

similar performance in both cases, particularly when $N = 120$. This observation underscores the capability of APUB-M to adjust to different uncertainty levels and the positive effect of increased sample sizes in lessening disparities caused by system noise.

3.5.3.2 Optimal Solutions Analysis.



(a) $N = 30$ (b) $N = 60$

(c) $N = 120$

Figure 3.7: Optimal order quantities of the ten products.

We now compare the optimal solutions of Models SAA-M and APUB-M in Case I, as illustrated in Figures 3.7a through 3.7c, with N varying from 30 to 120. These solutions dictate the recommended order quantities for the ten products.

When $N = 30$, Figure 3.7a depicts significant fluctuations in the curves of the order quantities as $(1-\alpha)$ increases, notably for product P2. The order quantity of product P2 decreases by 10.14% from SAA-M to APUB-M with $(1-\alpha) = 0.5$. The increase in N noticeably stabilizes the order quantities. Upon reaching $N = 60$, the relative difference in the order quantities of product P2 reduces to 9.95% between SAA-M and APUB-M with $(1-\alpha) = 0.5$. With a larger $N = 120$, all curves become flattened. In this scenario, as depicted in Figure 3.6c, APUB-M and SAA-M appear to achieve comparable performance. It can be seen in Figure 3.7c that their recommended optimal solutions are also very close.

SAA-M appears to be much more sensitive to N than APUB-M. Let us quantify the difference between the two solutions using the 2-norm. When N changes from 30 to 120 , the difference is 7.89 for SAA-M, 3.99 for APUB-M with $(1-\alpha) = 0.5$, and 3.36 for APUB-M with $(1-\alpha) = 0.95$. This observation suggests that APUB-M can provide a high-quality optimal solution even with a small sample size. This capability underscores the ability of APUB-M to simulate scenarios typically requiring a larger volume of data.

3.6 Conclusions

In this work, we introduce APUB, a novel statistical upper bound that acts as a critical bridge between the realms of statistical upper bounds and stochastic optimization. APUB enriches the theoretical landscape and highlights practical implications for the interpretability and

application of stochastic optimization models. APUB serves as both an upper bound for the population mean, enhancing statistical analysis, and a coherent risk measure for the sample mean, focusing particularly on tail distribution errors due to insufficient sample sizes. By rigorously proving the statistical soundness of our approach, including its asymptotic correctness and consistency, we lay a solid foundation for integrating statistical methods into decision-making frameworks under distributional ambiguity.

Furthermore, APUB-M that we innovatively develop integrates APUB into stochastic optimization. This integration makes the reliability of APUB-M transparent, by ensuring that the attribute of asymptotic correctness inherent in APUB is seamlessly transferred to APUB-M in the optimization context. Indeed, the coverage probability of APUB-M aligns with the concept of first-order correctness at this predefined nominal confidence level. Also, we meticulously show the asymptotic consistency of APUB-M, ensuring that our approach remains the nature of data-driven statistical methods, thereby promising stability and unbiasedness of APUB-M, as the sample size increases, but avoiding over-conservatism.

We employ a bootstrap sampling approximation method, BP-APUB-M, to manage the computational complexity, demonstrating that a significantly smaller number of bootstrap samples effectively maintains model integrity and reliability. This solution approach, underpinning the practical viability of APUB-M, confirms its applicability in real-world scenarios, particularly in two-stage linear stochastic optimization with random recourse. Moreover, our empirical studies across various stochastic optimization problems, including single-stage and two-stage models, underscore the robustness and practicality of APUB-M. The comparative analysis with traditional DRO methods, particularly in settings of fixed and random recourse, highlights the enhanced interpretability and reduced conservatism of APUB-M. These results not only validate our

theoretical findings but also showcase the broad applicability and effectiveness of our approach in real-world scenarios.

Appendix A Theorems Used in Our Proofs

Theorem A. 1 (Glivenko-Cantelli Theorem). Suppose that $X_1, X_2, \cdots$, *are independent and have a common distribution function* $\mathcal{F}$. *Denote by* $\mathcal{F}_N$ *the empirical cumulative distribution function Then,*

$$\sup_x \left|\mathcal{F}_N(x) - \mathcal{F}(x)\right| \to 0 \text{w.p.1.}$$

Proof. The proof is provided in (Billingsley, 2017, Theorem 20.6)

Theorem A.2 (Theorem 1, Rockafellar and Uryasev (2000)). *Let* $h(x,\omega)$ *be a random function where* $x \in \mathcal{X}$ *and* ω *belongs to an arbitrary probability space with distribution* Q. *Let* $q_\alpha(x)$ *denote the* $100(1-\alpha)$*-percentile of* $h(x,\omega)$ *and*

$$H_\alpha(x,t) = t + \frac{1}{\alpha}\int [h(x,\omega) - t]_+ \mathbb{Q}(d\omega)$$

where $t \in \mathbb{R}$. *Then, for all* $x \in \mathcal{X}$, *we have*

$$\frac{1}{\alpha}\int_0^\alpha q_\tau(x)\, d\tau = \min_{t\in\mathbb{R}} H_\alpha(x,t)$$

Theorem A.3 *Efron's percentile upper bound is 1st-order accurate.*

Proof. Efron (1981, Section 4) first proposed this method. He considers this bound as the limit of bootstrap percentile when infinite bootstrap samples are taken. The formal proof can be found in Section 4.2 Shao and Tu (2012).

Theorem A.4 (Theorem 2, Athreya (1983)). Suppose $\liminf MN^{-\phi} > 0$ for some $\phi > 0$ as $M, N \to \infty$, and $\mathbb{E}_{\mathbb{P}} |F(\xi) - \mu|^{\theta} < \infty$ for some $\theta \geq 1$ such that $\theta\phi > 1$. Then, as $M, N \to \infty$, we have

$$\frac{1}{M}\sum_{m=1}^{M} F\left(\zeta_m\left(\hat{\mathbb{P}}_N\right)\right) \to 1 \text{ w.p.1}$$

Theorem A.5 (Lemma 21.2, Vaart (1998)). *The quantile function of a cumulative distribution function* $\mathcal{F}$ *is the generalized inverse* $\mathcal{F}^{-1}: (0,1) \to \mathbb{R}$ *given by*

$$\mathcal{F}^{-1}(p) = \inf\left\{x : \mathcal{F}(x) \leq p\right\}$$

For any any sequence of cumulative distribution functions, $\mathcal{F}_N$ *converges to* $\mathcal{F}$ *in distribution if and only if* $\mathcal{F}_N^{-1}$ *converges to* $\mathcal{F}^{-1}$ *in distribution.*

Theorem A.6 (Theorem 2.1, Mallows and Richter (1969)). *Let* η *be a random variable,* $\mathcal{A}$ *be an event, and* σ_η *be its standard deviation. Then we have:*

$$\left|\mathbb{E}[\eta | \mathcal{A}] - \mathbb{E}[\eta]\right| \leq \sigma_\eta \left(\frac{1 - \Pr(\mathcal{A})}{\Pr(\mathcal{A})}\right)^{1/2}$$

Theorem A.7 (Theorem 10.8 (Rockafellar, 2015)). Let $\mathcal{C}$ be an open convex set. *Let* $(g_1, g_2, \cdots)$ *be a sequence of finite convex functions on* $\mathcal{C}$*. Suppose that the sequence converges pointwise on a dense subset* $\mathcal{D} \subseteq \mathcal{C}$ *and the limit is finite. Then, the sequence* $(g_1, g_2, \cdots)$ *converges uniformly to a continuous function on any compact subset inside* $\mathcal{C}$*.*

Theorem A.8 (Arzelà-Ascoli's Theorem). *Consider a sequence of real-valued continuous functions* $\{f_n\}_{n \in \mathbb{N}}$ *defined on a closed and bounded interval* $[a, b]$ *of the real line. If this sequence*

is uniformly bounded and uniformly equicontinuous, then there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ *that converges uniformly.*

Proof. The proof is well-known and can be found in Section 10.1 by Royden and Fitzpatrick [1968].

Theorem A.9 (Theorem 7.53, Shapiro et al. (2021)). *Let* $\mathcal{K}$ *be a nonempty compact subset of* $\mathbb{R}^n$ *and suppose that (i) for any* $x\in\mathcal{K}$ *the function* $F(\cdot,\xi)$ *is continuous at* x *for almost every* $\xi\in\Xi$*, (ii)* $F(x,\xi), x\in\mathcal{K}$*, is dominated by an integrable function, and (iii) the sample is iid. Then, the expected function* $f(x)$ *is finite valued and continuous on* $\mathcal{K}$*, and the sample mean* $\hat{f}_N(x)$ *converges to* $f(x)$ *w.p. 1 uniformly on* $\mathcal{K}$*.*

Theorem A.10 (Theorem 5.3, Shapiro et al. (2021)). *Suppose that there exists a compact set* $\mathcal{K}\subseteq\mathbb{R}^n$ *such that (i) the set* $\mathcal{S}$ *of the optimal solutions of the true problem is nonempty and is contained in* $\mathcal{K}$*, (ii) the function* $f(x)$ *is finite valued and continuous on* $\mathcal{K}$*, (iii) the sample average* $\hat{f}_N(x)$ *converges* $f(x)$ *w.p.1, as* $N\to\infty$*, uniformly in* $x\in\mathcal{K}$*, and (iv) w.p. 1 for* N *large enough the set* $\hat{\mathcal{S}}_N$ *is nonempty and contained in* $\mathcal{K}$*. Then, optimal values and solution set converges w.p.1 as* $N\to\infty$*.*

Theorem A.11 (Theorem 9.7, Birge and Louveaux (2011)). Consider a function $h:\mathcal{X}\times\Xi\to\mathbb{R}$. *Denote*

$$H(x)=\mathbb{E}\left[h(x,\xi)\right]$$

and

$$H_N(x)=\frac{1}{N}\sum_{i=1}^{N}h(x,\xi_i)$$

where $(\xi_1,\ldots,\xi_N)$ *are i.i.d. sample. Let* x^* *solve*

$$\min_x H(x)$$

and x_N^* *solve*

$$\min_x H_N(x)$$

Suppose there exist $a>0,\theta_0>0,\eta:\Xi\to\mathbb{R}$ *such that*

$|h(x,\xi)|\le a\eta(\xi)$ *and* $\mathbb{E}\left[e^{\theta\eta(\xi)}\right]<\infty$

for all $x\in\mathcal{X}$ *and for all* $0\le\theta\le\theta_0$. *Then, for any* ε, *there are* $\mathfrak{a}_\varepsilon,\mathfrak{b}_\varepsilon$ *such that*

$$\Pr\left(H\left(x_N^*\right)-H\left(x^*\right)\ge\varepsilon\right)\le\mathfrak{a}_\varepsilon e^{-\mathfrak{b}_\varepsilon N}$$

for all $N>0$. *If* x^* *is unique, then there are* $\mathfrak{c}_\varepsilon,\mathfrak{d}_\varepsilon$ *such that*

$$\Pr\left(x_N^*-x^*\ge\varepsilon\right)\le\mathfrak{c}_\varepsilon e^{-\mathfrak{d}_\varepsilon N}$$

for all $N>1$.

Appendix B Data for the Newsvendor problem in Section 3.5.3

$$p = -2, h = 9, b = 5$$

$$\mu_1 = [60.89, 48.58, 46.81, 56.54, 61.58, 52.69, 69.42, 60.54, 54.43, 51.76]^\top$$

$$\mu_2 = [50.30, 61.87, 53.16, 41.79, 51.94, 62.14, 45.47, 45.26, 55.95, 55.95]^\top$$

$$\Sigma_1 = \begin{bmatrix} 9.27 & 2.84 & -0.07 & 1.19 & -0.48 & 1.40 & 2.87 & 4.06 & -1.40 & -1.96 \\ 2.84 & 5.90 & -2.83 & 0.21 & 2.27 & -2.40 & -0.89 & 4.22 & 3.43 & 2.78 \\ -0.07 & -2.83 & 5.48 & -0.30 & 0.90 & 3.54 & -4.51 & -2.45 & -2.91 & -4.95 \\ 1.19 & 0.21 & -0.30 & 7.99 & -1.02 & -1.27 & -0.15 & -1.55 & -1.69 & -0.36 \\ -0.48 & 2.27 & 0.90 & -1.02 & 9.48 & -0.08 & -3.69 & 2.71 & -0.69 & -0.34 \\ 1.40 & -2.40 & 3.54 & -1.27 & -0.08 & 6.94 & -1.26 & -2.73 & 0.01 & -5.19 \\ 2.87 & -0.89 & -4.51 & -0.15 & -3.69 & -1.26 & 12.05 & -0.16 & -0.16 & 2.44 \\ 4.06 & 4.22 & -2.45 & -1.55 & 2.71 & -2.73 & -0.16 & 9.16 & -0.77 & 1.94 \\ -1.40 & 3.43 & -2.91 & -1.69 & -0.69 & 0.01 & -0.16 & -0.77 & 7.41 & 2.24 \\ -1.96 & 2.78 & -4.95 & -0.36 & -0.34 & -5.19 & 2.44 & 1.94 & 2.24 & 6.70 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 6.32 & 2.99 & -0.06 & 0.73 & -0.33 & 1.36 & 1.55 & 2.51 & -1.19 & -1.75 \\ 2.99 & 9.57 & -4.09 & 0.19 & 2.44 & -3.60 & -0.74 & 4.02 & 4.49 & 3.83 \\ -0.06 & -4.09 & 7.06 & -0.25 & 0.86 & 4.74 & -3.35 & -2.08 & -3.40 & -6.08 \\ 0.73 & 0.19 & -0.25 & 4.37 & -0.64 & -1.11 & -0.07 & -0.86 & -1.29 & -0.29 \\ -0.33 & 2.44 & 0.86 & -0.64 & 6.74 & -0.08 & -2.04 & 1.71 & -0.60 & -0.31 \\ 1.36 & -3.60 & 4.74 & -1.11 & -0.08 & 9.65 & -0.98 & -2.41 & 0.01 & -6.62 \\ 1.55 & -0.74 & -3.35 & -0.07 & -2.04 & -0.98 & 5.17 & -0.08 & -0.10 & 1.72 \\ 2.51 & 4.02 & -2.08 & -0.86 & 1.71 & -2.41 & -0.08 & 5.12 & -0.59 & 1.57 \\ -1.19 & 4.49 & -3.40 & -1.29 & -0.60 & 0.01 & -0.10 & -0.59 & 7.83 & 2.49 \\ -1.75 & 3.83 & -6.08 & -0.29 & -0.31 & -6.62 & 1.72 & 1.57 & 2.49 & 7.83 \end{bmatrix}$$

$$\varepsilon \sim \begin{bmatrix} \mathcal{U}(-5.37, 26.27) \\ \mathcal{U}(6.74, 14.16) \\ \mathcal{U}(3.22, 17.68) \\ \mathcal{U}(-7.48, 28.38) \\ \mathcal{U}(-4.89, 25.79) \\ \mathcal{U}(-0.21, 16.11) \\ \mathcal{U}(-12.14, 32.99) \\ \mathcal{U}(-7.74, 28.64) \\ \mathcal{U}(0.77, 20.13) \\ \mathcal{U}(2.13, 18.77) \end{bmatrix}$$

References

Abbasi-yadkori. Y, Pál. D, and Szepesvári. C. (2011) Improved Algorithms for Linear Stochastic Bandits. In *Advances in Neural Information Processing Systems,* 24. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper_files/paper/2011/hash/e1d5be1c7f2f456670de3d53c7b54f4a-Abstract.html.

Arias. N. B, Tabares. A, Franco. J. F, Lavorato. M, and Romero. R. (2017) Robust joint expansion planning of electrical distribution systems and ev charging stations. *IEEE Transactions on Sustainable Energy,* 9(2):884-894.

Askari. B, Le Quy. T, and Ntoutsi. E. (July 2020) Taxi Demand Prediction using an LSTM-Based Deep Sequence Model and Points of Interest. In *2020 IEEE 44 th Annual Computers, Software, and Applications Conference (COMPSAC),* pages 1719-1724. doi: 10.1109/COMPSAC48688.2020.000-7. URL https://ieeexplore.ieee.org/document/9202791.

Athreya. K. B. (Mar. 1983) Strong law for the bootstrap. *Statistics & Probability Letters,* 1(3): 147-150. ISSN 0167-7152. doi: 10.1016/0167-7152(83)90063-9. URL https://www.sciencedirect.com/science/article/pii/0167715283900639.

Auer. P, N. Cesa-Bianchi, and Fischer. P. (May 2002) Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning,* 47(2):235-256. ISSN 1573-0565. doi: 10.1023/A:1013689704352. URL https://doi.org/10.1023/A:1013689704352.

Baran. M. E. and Wu. F. F. (1989) Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Power Engineering Review,* 9(4):101-102.

Bauer. G. S., Zheng. C., Shaheen. S., and Kammen. D. M. (2021) Leveraging big data and coordinated charging for effective taxi fleet electrification: The 100% ev conversion of shenzhen, china. *IEEE Transactions on Intelligent Transportation Systems,* 23 (8):10343-10353.

Bayraksan. G. and Love. D. K. (Sept. 2015) Data-Driven Stochastic Programming Using Phi Divergences. In *The Operations Research Revolution,* INFORMS TutORials in Operations Research, pages 1-19. INFORMS. ISBN 978-0-9843378-8-0. doi: 10.1287/educ.2015.0134. URL https://pubsonline.informs.org/doi/abs/ 10.1287/educ.2015.0134. Section: 1.

Ben-Tal A., den Hertog D., De Waegenaere A., Melenberg B., and Rennen. G. (Feb. 2013) Robust Solutions of Optimization Problems Affected by Uncertain Probabilities. *Management Science,* 59(2):341-357. ISSN 0025-1909. doi: 10.1287/mnsc.1120.1641. URL https://pubsonline.informs.org/doi/abs/10. 1287/mnsc.1120.1641. Publisher: INFORMS.

Billingsley. P. (2017) *Probability and measure.* John Wiley & Sons. ISBN 978-1-11862596-5. Google-Books-ID: 6ItqtwaWZZQC.

Birge J. R. and Louveaux. F. (June 2011) *Introduction to Stochastic Programming.* Springer Science & Business Media. ISBN 978-1-4614-0237-4. Google-Books-ID: Vp0Bp8kjPxUC.

Blanchet J. and Murthy. K. (May 2019) Quantifying Distributional Model Risk via Optimal Transport. *Mathematics of Operations Research,* 44(2):565-600. ISSN 0364765X. doi: 10.1287/moor.2018.0936. URL https://pubsonline.informs.org/ doi/abs/10.1287/moor.2018.0936. Publisher: INFORMS.

Cai H., Jia X., Chiu A. S., Hu X., and Xu. M. (2014) Siting public electric vehicle charging stations in beijing using big-data informed travel patterns of the taxi fleet. *Transportation Research Part D: Transport and Environment,* 33:39-46.

Calafiore G. C. and Ghaoui. L. E. (July 2006) On Distributionally Robust Chance-Constrained Linear Programs. *Journal of Optimization Theory and Applications,* 130(1):1-22. ISSN 1573-2878. doi: 10.1007/s10957-006-9084-x. URL https://doi. org/10.1007/s10957-006-9084-x.

Chung H.-M., Li W.-T., Yuen C., Wen C.-K., and Crespi. N. (2018) Electric vehicle charge scheduling mechanism to maximize cost efficiency and user convenience. *IEEE Transactions on Smart Grid,* 10(3):3020-3030.

Cilio L. and Babacan. O. (2021) Allocation optimisation of rapid charging stations in large urban areas to support fully electric taxi fleets. *Applied Energy*, 295:117072.

Clairand J.-M., González-Rodríguez M., Kumar R., Vyas S., and Escrivá-Escrivá. G. (2022) Optimal siting and sizing of electric taxi charging stations considering transportation and power system requirements. *Energy,* 256:124572.

Dantzig. G. B. (Aug. 2016) Linear Programming and Extensions. In *Linear Programming and Extensions.* Princeton University Press. ISBN 978-1-4008-8417-9. doi: 10.1515/9781400884179. URL https://www.degruyter.com/document/doi/10. 1515/9781400884179/html.

Quevedo P. M. de, Munoz-Delgado G., and Contreras. J. (2017) Impact of electric vehicles on the expansion planning of distribution systems considering renewable energy, storage, and charging stations. *IEEE Transactions on Smart Grid,* 10(1):794-804.

Deb S., Tammi K., Gao X.-Z., Kalita K., Mahanta P., and Cross. S. (2021) A robust twostage planning model for the charging station placement problem considering road traffic uncertainty. *IEEE Transactions on Intelligent Transportation Systems,* 23(7):6571-6585.

Delage E. and Ye. Y. (June 2010) Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems. *Operations Research,* 58(3):595612. ISSN 0030-364X. doi: 10.1287/opre.1090.0741. URL https: //pubsonline.informs.org/doi/abs/10.1287/opre.1090.0741. Publisher: INFORMS.

Devore. J. (Mar. 2009) Probability and statistics for engineering and the sciences. URL https://hero.epa.gov/hero/index.cfm/reference/details/reference_ id/196740.

Drèze J. H. and Rustichini. A. (2004) *Handbook of utility theory,* volume 2, chapter Statedependent utility and decision theory, pages 839-892. Kluwer Academic Publishers, Boston, MA.

DSA Subcommitte. IEEE 123 node test feeder. Accessed: March. 7, 2023. [Online]. Available: https://cmte.ieee.org/pes-testfeeders/resources/.

Duque D., Mehrotra S., and Morton. D. P. (2022) Distributionally robust two-stage stochastic *programming. SIAM Journal on Optimization,* 32(3):1499-1522.

Efron. B. (1981) Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics,* 9(2):139-158. ISSN 1708-945X. doi: 10.2307/3314608. URL https://onlinelibrary.wiley.com/doi/abs/10.2307/3314608. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/3314608.

Efron. B. (1987) Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association,* 82(397):171-185. ISSN 0162-1459. doi: 10.2307/2289144. URL

https://www.jstor.org/stable/2289144. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Egbue O. and Long. S. (2012) Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy policy*, 48:717-729.

Ehsan A. and Yang. Q. (2019) Active distribution system reinforcement planning with ev charging stations - part i: Uncertainty modeling and problem formulation. *IEEE Transactions on Sustainable Energy,* 11(2):970-978.

El-Hawary. M. E. (2008) *Introduction to electrical power systems,* volume 50. John Wiley & Sons.

Gangrade A., P. Pratyush, and G. Hajela. (2022) Taxi-demand forecasting using dynamic spatiotemporal analysis. *ETRI Journal,* 44(4):624-640.

Gao R. and Kleywegt. A. (May 2023) Distributionally Robust Stochastic Optimization with Wasserstein Distance. *Mathematics of Operations Research,* 48(2):603-655. ISSN 0364-765X. doi: 10.1287/moor.2022.1275. URL https://pubsonline. informs.org/doi/full/10.1287/moor.2022.1275. Publisher: INFORMS.

Greenshields B. D., Bibbins J., Channing W., and Miller. H. (1935) A study of traffic capacity. In *Highway research board proceedings,* volume 14, pages 448-477. Washington, DC.

Guan Z., Xiang Z., Shen B., Bao X., and Bojun. C. (2020) Electrify taxis in new york city. [Online; accessed 22-Sep-2021].

Guo Z., Zhou Z., and Zhou. Y. (2019) Impacts of integrating topology reconfiguration and vehicle-to-grid technologies on distribution system operation. *IEEE Transactions on Sustainable Energy,* 11(2):1023-1032.

Hagman J. and Langbroek. J. H. (2019) Conditions for electric vehicle taxi: A case study in the greater stockholm region. *International Journal of Sustainable Transportation,* 13(6):450-459.

Hall. P. (1986) On the Bootstrap and Confidence Intervals. *The Annals of Statistics,* 14(4): 1431-1452. ISSN 0090-5364. URL https://www.jstor.org/stable/2241480. Publisher: Institute of Mathematical Statistics.

Hanasusanto G. A., D. Kuhn, S. W. Wallace, and S. Zymler. (Aug. 2015) Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming,* 152(1):1-32. ISSN 1436-4646. doi: 10.1007/ s10107-014-0776-y. URL https://doi.org/10.1007/s10107-014-0776-y.

Hazra. A. (Oct. 2017) Using the confidence interval confidently. *Journal of Thoracic Disease,* 9 (10):4125-4130. ISSN 2072-1439. doi: 10.21037/jtd.2017.09.14. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5723800/.

Hirschfeld. A. (2019) Taxis go electric: Tesla joins cab fleets in new york city. Accessed: Oct. 10, 2021. [Online]. Available: https://observer.com/2019/11/ tesla-taxi-electric-vehicles-cab-fleets/.

House. T. W. (2021) Biden administration advances electric vehicle charging infrastructure. URL https://www. whitehouse.gov/briefing-room/statements-releases/2021/04/22/ fact-sheet-biden-administration-advances-electric-vehicle-charging-infrastructure/.

Huang S., Kuo-Ling; Mehrotra. (2016) iOptimize. Accessed: Mar. 15, 2021. [Online]. Available :https://www.swmath.org/software/20377. Indeed. Taxi driver salary in new york, ny, 2023. Accessed: Oct. 01, 2023. [Online]. Available at:https://www.indeed.com/career/taxi-driver/salaries/ New-York--NY.

Kaya Ö., Alemdar K. D., and Çodur. M. Y. (2020) A novel two stage approach for electric taxis charging station site selection. *Sustainable Cities and Society,* 62:102396.

Laporte G. and Louveaux. F. V. (1993) The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters,* 13(3):133-142.

Leblanc. L. J. (1975) An algorithm for the discrete network design problem. *Transportation Science,* 9(3):183-199.

Li M., Ye H., Liao X., Ji J., and Ma. X. (2020) How shenzhen, china pioneered the widespread adoption of electric vehicles in a major city: Implications for global implementation. *Wiley Interdisciplinary Reviews: Energy and Environment,* 9(4):e373.

Lin F., Fang X., and Gao. Z. (2022) Distributionally robust optimization: A review on theory and applications. *Numerical Algebra,* Control and Optimization, 12(1):159 212.

Malhotra A., Binetti G., Davoudi A., and Schizas. I. D. (2016) Distributed power profile tracking for heterogeneous charging of electric vehicles. *IEEE Transactions on Smart Grid,* 8(5):2090-2099.

Mallows C. L. and Richter. D. (1969) Inequalities of Chebyshev Type Involving Conditional Expectations. *The Annals of Mathematical Statistics*, 40(6):1922-1932. ISSN 0003-4851. URL https://www.jstor.org/stable/2239511. Publisher: Institute of Mathematical Statistics.

Martinez C. M., Hu X., Cao D., Velenis E., Gao B., and Wellers. M. (2016) Energy management in plug-in hybrid electric vehicles: Recent progress and a connected vehicles perspective. *IEEE Transactions on Vehicular Technology*, 66(6):4534-4549.

Meng X., Zhang W., Bao Y., Yan Y., Yuan R., Chen Z., and Li. J. (2020) Sequential construction planning of electric taxi charging stations considering the development of charging demand. *Journal of Cleaner Production,* 259:120794.

Mnih V., Szepesvári C., and Audibert. J.-Y. (July 2008) Empirical Bernstein stopping. In *Proceedings of the 25th international conference on Machine learning*, ICML '08, pages 672-679, New York, NY, USA. Association for Computing Machinery. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390241. URL https: //dl.acm.org/doi/10.1145/1390156.1390241.

Mohajerin Esfahani P. and Kuhn. D. (Sept. 2018) Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115-166. ISSN 1436-4646. doi: 10. 1007/s10107-017-1172-1. URL https://doi.org/10.1007/s10107-017-1172-1.

Nelles. Hughes H., (1984) " networks of power: Electrification in western society, 18801930"(book review). In *Canadian Journal of History/Annales Canadiennes d' Histoire*, volume 19, page 158. Journal of History Co.

NYC TAXI LIMOUSINE COMMISSION. (2013) Take charge a roadmap to electric new york city taxis. Accessed: Feb. 15, 2020 [Online]. Available: https://www1.nyc.gov/assets/tlc/downloads/pdf/electric_taxi_ task_force_report_20131231.pdf.

PJM. Data miner 2, (2022). Accessed: Mar. 15, 2022. [Online]. Available at:https: //dataminer2.pjm.com/list.

Plautz. J. (2018) Columbus, oh puts up $30k for electric taxis. Accessed: Oct. 10, 2021. [Online]. Available at: https://www.smartcitiesdive.com/news/ columbus-ohio-electric-taxis/525995/.

Rahimian H. and Mehrotra. S. (2019) Distributionally robust optimization: A review. *ar X iv preprint arXiv:1908.*05659.

Read T. R. C. and Cressie. N. (2012) Goodness-Of-Fit Statistics for Discrete Multivariate Data | Semantic Scholar. URL https://www.semanticscholar. org/paper/Goodness-Of-Fit-Statistics-for-Discrete-Data-Read-Cressie/ 563e27a8b8c4b1e412f306be93cea9467789b560.

Rockafellar. R. T. (Apr. 2015) Convex Analysis: (PMS-28). In Convex Analysis. Princeton University Press. ISBN 978-1-4008-7317-3. doi: 10.1515/ 9781400873173. URL https://www.degruyter.com/document/doi/10.1515/ 9781400873173/ html?lang=en.

Rockafellar R. T. and Uryasev. S. (2000) Optimization of conditional value-at-risk. *The Journal of Risk,* 2(3):21-41. ISSN 14651211. doi: 10.21314/JOR.2000. 038. URL http://www.risk.net/journal-of-risk/technical-paper/2161159/ optimization-conditional-value-risk.

Rodrigues F., Markou I., and Pereira F. C.. (Sept, 2019) Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach. *Information Fusion,* 49:120-129. ISSN 1566-2535. doi: 10.1016/j.inffus.2018.07.007.

Royden H. L. and Fitzpatrick. P. (1968) *Real analysis,* volume 2. Macmillan New York. URL https://cir.nii.ac.jp/crid/1130282270315169024.

Sarykalin S., Serraino G., and Uryasev. S. (Sept. 2008) Value-at-Risk vs. Conditional Value-at Risk in Risk Management and Optimization. In *State-of-the-Art Decision-Making Tools in the Information-Intensive Age,* INFORMS TutORials in Operations Research, pages 270-294. INFORMS. ISBN 978-1-877640-23-0. doi: 10.1287/educ.1080.0052. URL https://pubsonline.informs.org/doi/abs/10. 1287/educ.1080.0052. Section: 13.

Shao J. and Tu. D. (Dec. 2012) *The Jackknife and Bootstrap.* Springer Science & Business Media. ISBN 978-1-4612-0795-5. Google-Books-ID: VO3SBwAAQBAJ.

Shapiro A., Dentcheva D., and Ruszczyński. A. (2009) *Lectures on Stochastic Programming: Modeling and Theory.* Society for Industrial and Applied Mathematics.

Shapiro A., Dentcheva D., and Ruszczynski. A. (July 2021) *Lectures on Stochastic Programming: Modeling and Theory, Third Edition.* MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics. ISBN 978-1-61197-658-8. doi: 10.1137/1.9781611976595. URL https://epubs.siam.org/doi/book/10.1137/1.9781611976595.

Smith J. E. and Winkler. R. L. (2006) The Optimizer's Curse: Skepticism and Postdecision Surprise in Decision Analysis. *Management Science,* 52(3):311-322. ISSN 00251909. URL https://www.jstor.org/stable/20110511. Publisher: INFORMS.

Sossan F., Mukherjee B., and Hu. Z. (2020) Impact of the charging demand of electric vehicles on distribution grids: a comparison between autonomous and non-autonomous driving. *In 2020 Fifteenth International Conference on Ecological Vehicles and Renewable Energies (EVER),* pages 1-6. IEEE.

The White House. (2021) President biden announces steps to drive american leadership forward on clean cars and trucks. Accessed: Dec. 11, 2021. [Online]. Available: https://www.whitehouse.gov/.

Vaart. A. W. v. d. (1998) *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. ISBN 978-0-52178450-4. doi: 10.1017/CBO9780511802256. URL https://www.cambridge.org/ core/books/asymptotic-statistics/A3C7DAD3F7E66A1FA60E9C8FE132EE1D.

Wang S., Bi S., Zhang Y.-J. A., and Huang. J. (2018) Electrical vehicle charging station profit maximization: Admission, pricing, and online scheduling. *IEEE Transactions on Sustainable Energy,* 9(4):1722-1731.2018a

Wang S., Dong Z. Y., Chen C., Fan H., and Luo. F. (2019) Expansion planning of active distribution networks with multiple distributed energy resources and ev sharing system. *IEEE Transactions on Smart Grid,* 11(1):602-611.

Wang X., Shahidehpour M., Jiang C., and Li. Z. (2018b) Coordinated planning strategy for electric vehicle charging stations and coupled traffic-electric networks. *IEEE Transactions on Power Systems,* 34(1):268-279.

Wei W., Danman W., Qiuwei W., Shafie-Khah M., and Catalao. J. P. (2019) Interdependence between transportation system and power distribution system: a comprehensive review on models and applications. *Journal of Modern Power Systems and Clean Energy,* 7(3):433-448.

Wen C.-K., Chen J.-C., Teng J.-H., and Ting. P. (2012) Decentralized plug-in electric vehicle charging selection algorithm in power systems. *IEEE Transactions on Smart Grid,* 3(4):1779-1789.

Wiesemann W., Kuhn D., and Sim. M. (Dec. 2014) Distributionally Robust Convex Optimization. *Operations Research,* 62(6):1358-1376. ISSN 0030-364X. doi: 10.1287/opre.2014.1314. URL https://pubsonline.informs.org/doi/10.1287/ opre.2014.1314. Publisher: INFORMS.

Xia M., Lai Q., Zhong Y., Li C., and Chiang. H.-D. (2016) Aggregator-based interactive charging management system for electric vehicle charging. *Energies,* 9(3):159.

Xie. W. (June 2020) Tractable reformulations of two-stage distributionally robust linear programs over the type- ∞ Wasserstein ball. *Operations Research Letters,* 48. doi: 10.1016/j.orl.2020.06.003.

Yang. W. (2018) A user-choice model for locating congested fast charging stations. *Transportation Research Part E: Logistics and Transportation Review,* 110:189-213.

Yang Z., Zhong H., Bose A., Zheng T., Xia Q., and Kang. C. (2017) A linearized opf model with reactive power and voltage magnitude: A pathway to improve the mw-only dc opf. *IEEE Transactions on Power Systems,* 33(2):1734-1745.

Yang Z., Xie K., Yu J., Zhong H., Zhang N., and Xia. Q. (2018) A general formulation of linear power flow models: Basic theory and error analysis. IEEE Transactions on Power Systems, 34(2):1315-1324.

Yao H., Wu F., Ke J., Tang X., Jia Y., Lu S., Gong P., Ye J., and Li. Z. (Feb. 2018) Deep Multi-View Spatial-Temporal Network for Taxi Demand Prediction. URL http://arxiv.org/abs/1802.08714.

Zhang J., Pourazarm S., Cassandras C. G., and Paschalidis. I. C. (2016) The price of anarchy in transportation networks by estimating user cost functions from actual traffic data. In *2016 IEEE 55th Conference on Decision and Control (CDC),* pages 789-794. IEEE.

Zhang S., Wang H., Zhang Y.-f., and Li. Y.-Z. (2019) A novel two-stage location model of charging station considering dynamic distribution of electric taxis. *Sustainable Cities and Society*, 51:101752.