Data-Driven Distributionally Robust Optimization for Long-Term Contract vs. Spot Allocation Decisions

Dimitri J. Papageorgiou [a, 1]

[a] ExxonMobil Technology and Engineering Company – Research Division, 1545 Route 22 East, Annandale, NJ, 08801

Abstract

There are numerous industrial settings in which a decision maker must decide whether to enter into long-term contracts to guarantee price (and hence cash flow) stability or to participate in more volatile spot markets. In this paper, we investigate a data-driven distributionally robust optimization (DRO) approach aimed at balancing this tradeoff. Unlike traditional risk-neutral stochastic optimization models that assume the underlying probability distribution generating the data is known, DRO models assume the distribution belongs to a family of possible distributions, thus providing a degree of immunization against unseen and potential worst-case outcomes. We compare and contrast the performance of a risk-neutral model, conditional value-at-risk formulation, and a Wasserstein distributionally robust model to demonstrate the potential benefits of a DRO approach for an "elasticity-aware" price-taking decision maker.

Keywords

Conditional value-at-risk, Data-driven distributionally robust optimization, Price elasticity, Wasserstein metric.

Introduction

In numerous energy markets, including electricity and natural gas, energy providers/sellers face the dilemma of deciding how to allocate their supply across various markets and over time. Throughout this paper, for convenience, we will use the motivating example of a generation company (Genco) who sells electricity. There are typically at least two major types of markets where Gencos can sell electricity: the spot market and the forward market for longer-term bilateral contracts market (Kirschen and Strbac, 2018). Here, the spot market refers to a public financial market where electricity is traded on a daily, hourly, or subhourly basis. Gencos deliver electricity immediately and buyers pay for it "on the spot." Such markets can be highly volatile as supply and demand variability can cause the market-clearing price to fluctuate dramatically over time. In contrast, forward markets allow for buyers and sellers to enter longer-term bilateral contracts to reduce price variability over a time span of interest. Among other benefits, forward markets allow market participants to hedge against uncertainty by offering price predictability, which in turn allows the Genco to better plan its cash flows and potentially secure more favorable financing from lenders. A power purchase agreement (PPA) is one such example of a bilateral contract where a Genco enters a long-term agreement to provide (a typically fixed amount of) energy at a fixed price over many time periods, e.g., one year. This paper investigates the problem of deciding how to allocate supply within these two types of markets.

Electricity applications investigating long-term vs. spot tradeoffs

Given the importance of balancing risk exposure and expected profits in the power sector, many researchers have studied how power producers should simultaneously optimize contractual involvement, spot allocation, and generation planning. Within the electricity sector, this joint problem is often referred to as "power portfolio optimization" and "integrated risk management" (Lorca and Prina, 2014). Although there are a host of contracts available, in this work we focus exclusively on fixed-price forward contracts; we do not consider other options and derivatives. A forward contract is an agreement to buy or sell a fixed amount of electricity at a given price over a fixed time horizon.

The importance of forward contracts within the electricity industry has been known for decades (Kaye et al., 1990). We therefore attempt here to highlight key papers that have employed a rigorous approach to address power portfolio optimization under uncertainty. Early papers employed more traditional risk-neutral stochastic programming techniques as a means to transcend a deterministic Markowitz mean-variance mindset (Kwon et al., 2006; Sen et al., 2006). Risk-averse models soon followed to manage downside exposure.

[1] Corresponding author. Email: `dimitri.j.papageorgiou@exxonmobil.com`.

Noteworthy papers that employ the conditional value-at-risk ($\mathbb{CVAR}$) metric include Conejo et al. (2008), Street et al. (2009), Pineda and Conejo (2012), Yau et al. (2011), Lorca and Prina (2014). For example, Street et al. (2009) investigate bidding strategies for risk-averse Gencos in a long-term forward contract auction. Fanzeres et al. (2014) examine contracting strategies for renewable generators using an interesting hybrid stochastic/robust optimization approach.

Our investigation most closely aligns with the study conducted by Lorca and Prina (2014), who pose a risk-averse stochastic linear program for a "medium-term" planning horizon. The qualifier "medium-term" is used to distinguish the problem from short-term scheduling problems over minutes, hours, or days, and from long-term applications that may involve capital-intensive generation investment decisions. We also consider a medium-term planning horizon of one year.

Why pursue distributionally robust optimization?

Over the past decade, distributionally robust optimization (DRO) has emerged as a powerful tool within the operations research and statistical learning communities, while also garnering attention within the process systems engineering community (see, e.g., Gao et al. (2019); Liu and Yuan (2021); Shang and You (2018)). Rahimian and Mehrotra (2019) survey fundamental DRO concepts and applications, in addition to relating it with robust optimization, risk aversion, chance-constrained optimization, and function regularization. Loosely speaking, DRO is well-suited to address data-driven optimization problems as it puts faith in the empirical data, but not too much.

Esfahani and Kuhn (2018) motivate DRO quite nicely. A traditional stochastic program attempts to solve the problem $\min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}^{\mathbb{P}}[h(\mathbf{x},\boldsymbol{\xi})]$, where the loss function $h:\mathbb{R}^n\times\mathbb{R}^m$ depends on both the decision vector $\mathbf{x}\in\mathbb{R}^n$ and the random vector $\boldsymbol{\xi}\in\mathbb{R}^m$ governed by the distribution $\mathbb{P}$. Unfortunately, as many practitioners have discovered, the true distribution $\mathbb{P}$ is rarely known precisely and must be inferred from data, physics, expert knowledge, and more. Optimizing a traditional stochastic program can then lead to solutions that are "overfitted to the data." Second, computing the expectation in a stochastic program for a fixed decision $\mathbf{x}$ can be computationally challenging in its own right as it may involve the evaluation of a multivariate integral.

To combat these challenges, DRO attempts to hedge the expected loss against a family $\mathcal{P}$ of distributions that include the true data-generating mechanism with high confidence (Chen and Paschalidis, 2018). Mathematically, DRO minimizes the expected loss over the worst-case distribution $\mathbb{Q}\in\mathcal{P}$ by solving

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}^{\mathbb{Q}}\left[h(\mathbf{x},\boldsymbol{\xi})\right]. \tag{1}$$

The family $\mathcal{P}$ is also known as an *ambiguity set* and minimizing the inner "sup" expectation is sometimes termed an *ambiguity-averse* (as opposed to "risk-averse") problem, which is why DRO is also known as *ambiguous stochastic optimization* (Rahimian and Mehrotra, 2019). DRO "bridges the gap between data and decision-making – statistics and optimization frameworks – to protect the decision-maker from the ambiguity in the underlying probability distribution" (Rahimian and Mehrotra, 2019).

Existing DRO approaches can be divided into two broad classes – moment-based and statistical distance-based – according to the way in which the ambiguity set $\mathcal{P}$ is constructed. Moment-based ambiguity sets postulate that the empirical data must belong to a distribution that satisfies certain moment (e.g., mean and variance) constraints (Delage and Ye, 2010). While such approaches often give rise to tractable formulations, they sometimes produce overly conservative solutions (Wang et al., 2016) and may not enjoy favorable asymptotic consistency or finite sample guarantees (Hanasusanto and Kuhn, 2018). On the other hand, statistical distance-based ambiguity sets require distributions that are stastically "close" to the empirical distribution. Popular choices of distance metrics include Kullback-Leibler divergence, ϕ-divergence, the Prokhorov metric, total variation, and more (Rahimian and Mehrotra, 2019). Due to several shortcomings in the aforementioned metrics (Gao and Kleywegt, 2016), the Wasserstein distance metric has garnered considerable attention in the past decade, within both the machine learning and optimization communities. It possesses favorable statistical guarantees, while also leading to tractable optimization formulations (Gao and Kleywegt, 2016; Esfahani and Kuhn, 2018; Rahimian and Mehrotra, 2019). It is for these reasons that we pursue the Wasserstein metric in this study.

Contributions

The contributions of this paper are:

1. In contrast to the prevailing simplistic price-taker models commonly found in the literature, we consider an "elasticity-aware" decision-maker. Consequently, the supplier can estimate the price elasticity due to her own supply to each spot market and behaves accordingly so as not to oversaturate a particular market and ultimately overly depress prices.

2. We present a data-driven DRO approach exploiting Wasserstein ambiguity sets and contrast it against a standard conditional value-at-risk approach. As data-driven techniques that bridge data science, machine learning, and optimization hold significant promise, we believe that this investigation is valuable for the process systems engineering community.

3. We provide numerical evidence that our risk-averse models are tractable for an interesting application in the real-time PJM electricity market. We also explore the tradeoff between supply allocation to long-term contracts vs. the spot market as a function of the decision maker's risk aversion.

Problem Statement

Suppose that a key decision maker's objective is to maximize profit by selling a commodity (e.g., electricity) in a set

$\mathcal{M}$ of market locations over a fixed planning horizon. Within each market location $m \in \mathcal{M}$, she has two options available: (1) enter into long-term fixed-price forward bilateral contracts with customers to ensure price predictability, or (2) sell to one or more spot markets at a potentially volatile market-clearing price. More formally, she must choose a long-term contract amount $x^{\min}_{mc} \in \mathbb{R}_+$ (a non-negative scalar) for each market location $m \in \mathcal{M}$ and each long-term contract $c \in \mathcal{C}_m$, which holds for the entire planning horizon (e.g., one month). The contracted amount $x^{\min}_{mc}$ determines the minimum and maximum amount of a particular product/commodity that must be sold at a fixed price to the associated contract holder in each time period (e.g., hour) $t \in \mathcal{T}$. Specifically, she can sell up to $x^{\min}_{mc} + X^+_{mc}$ to long-term contracts where $X^+_{mc} \in \mathbb{R}_+$. Any remaining production in that time period can be sold to one or more spot markets. Let W_{mct} be the long-term price (sometimes called a "wholesale" price) associated with contract $c \in \mathcal{C}_m$ in market m in time period $t \in \mathcal{T}$. Oftentimes, the long-term contract price is not a function of time and can therefore be written simply as W_{mc}, but we will keep the subscript t for the more general setting in which the price is known to vary with time.

Meanwhile, we assume that spot market prices in each market location m are unknown when deciding long-term contract amounts $x^{\min}_{mc}$. In a data-driven setting, we assume that we have access to a finite set $\mathcal{S}$ of scenarios, e.g., a set of historical spot price time series. Associated with each scenario $s \in \mathcal{S}$ is a probability $\pi_s \in (0,1]$ and spot price curve for every market m and time period $t \in \mathcal{T}$. Naturally, $\sum_{s\in\mathcal{S}} \pi_s = 1$. Unlike the majority of price-taker models in the literature, we assume that the decision maker is an "elasticity-aware" price taking supplier, i.e., the supplier can estimate the price elasticity due to her own supply to each spot market m. We do not assume price setter behavior, however, in which suppliers could exert market power and could therefore act as Nash-Cournot players. In our first simplified setting, we assume that there is no cross-market price elasticity in which case all markets are independent. That is, the stochastic spot price in market location m_1 is independent of the price (and quantity sold) in market location m_2 for all $m_1 \neq m_2 \in \mathcal{M}$. This simplification allows us to represent the spot market price in each location via a descending "staircase" structure defined by a set $\mathcal{K}_m$ of steps. The height of step k in time period t in scenario s is denoted by P_{mkts} and satisfies $P_{m1ts} > P_{m2ts} > \cdots > P_{mKts}$ (with $K = |\mathcal{K}_m|$), and has width $Y^{\text{Spot}}_{mkts} > 0$. This spot price structure implies that at most Y^{Spot}_{m1ts} units can be sold to spot at price P_{m1ts} before the price decreases to P_{m2ts} and so forth. We then describe an extension to handle "separable" cross-market price elasticity.

We assume that the decision maker also has production and transportation costs to consider. Analogous to the aforementioned demand curves, we assume a standard supply curve represented by an increasing "staircase" structure. That is, the supplier can produce up to U^{Prod}_i units at a production cost of C^{Prod}_i for each supply step $i \in I$. Let L_t and U_t be the known minimum and maximum production limits in time period t. Let C^{Trans}_m denote the per-unit transportation cost to market location m. Our basic model implicitly assumes that all supply is co-located, e.g., at a centralized location; this assumption could easily be relaxed.

As for the decision variables, let x^{Term}_{mcts} and $\sum_{k\in\mathcal{K}_m} y^{\text{Spot}}_{mkts}$ denote the amount of production to allocate to long-term contract $c \in \mathcal{C}_m$ and to the spot market in location m, respectively, in time period t in scenario s. As stated above, $x^{\min}_{mc}$ denotes the long-term contract volume allocation to market-contract pair (m,c). Finally, u^{Prod}_{its} and u^{Trans}_{mts} denote the production amount at supply step $i \in I$ and the amount transported to market m in time period t in scenario s.

Risk-neutral stochastic programming formulation

Assuming no cross-market price elasticity, a potential scenario-based risk-neutral stochastic mixed-integer linear program (MILP) has the following form:

$$\max_{\substack{\mathbf{x}^{\min},\\ \mathbf{u},\mathbf{x},\mathbf{y},\mathbf{z}}} \quad \sum_{s\in\mathcal{S}} \pi_s z_s \quad (= \text{Expected Profit}) \tag{2a}$$

$$\text{s.t.} \quad z_s = \sum_{t\in\mathcal{T}} \Bigg[\sum_{m\in\mathcal{M}} \sum_{c\in\mathcal{C}_m} W_{mct} x^{\text{Term}}_{mcts} + \sum_{k\in\mathcal{K}_m} P_{mkts} y^{\text{Spot}}_{mkts} - \sum_{i\in I} C^{\text{Prod}}_i u^{\text{Prod}}_{its} - \sum_{m\in\mathcal{M}} C^{\text{Trans}}_m u^{\text{Trans}}_{mts} \Bigg] \quad \forall s \in \mathcal{S} \tag{2b}$$

$$x^{\min}_{mc} \leq x^{\text{Term}}_{mcts} \leq x^{\min}_{mc} + X^+_{mc} \quad \forall m, c \in \mathcal{C}_m, t, s \tag{2c}$$

$$\sum_{i\in I} u^{\text{Prod}}_{its} = \sum_{m\in\mathcal{M}} \sum_{c\in\mathcal{C}_m} x^{\text{Term}}_{mcts} + \sum_{k\in\mathcal{K}_m} y^{\text{Spot}}_{mkts} \quad \forall t \in \mathcal{T}, s \tag{2d}$$

$$\sum_{i\in I} u^{\text{Prod}}_{its} = \sum_{m\in\mathcal{M}} u^{\text{Trans}}_{mts} \quad \forall t \in \mathcal{T}, s \in \mathcal{S} \tag{2e}$$

$$L_t \leq \sum_{i\in I} u^{\text{Prod}}_{its} \leq U_t \quad \forall t \in \mathcal{T}, s \in \mathcal{S} \tag{2f}$$

$$(\mathbf{u}, \mathbf{x}^{\min}, \mathbf{x}^{\text{Term}}, \mathbf{y}^{\text{Spot}}, \mathbf{z}) \in \mathcal{X} \tag{2g}$$

$$u^{\text{Prod}}_{its} \in [0, U^{\text{Prod}}_i] \quad \forall i \in I, t \in \mathcal{T}, s \in \mathcal{S} \tag{2h}$$

$$x^{\min}_{mc} \in [0, X^{\max}_{mc}] \quad \forall m \in \mathcal{M}, c \in \mathcal{C}_m \tag{2i}$$

$$x^{\text{Term}}_{mcts} \in [0, X^{\max}_{mc}] \quad \forall m \in \mathcal{M}, c \in \mathcal{C}_m, t \in \mathcal{T}, s \tag{2j}$$

$$y^{\text{Spot}}_{mkts} \in [0, Y^{\text{Spot}}_{mkts}] \quad \forall m \in \mathcal{M}, k \in \mathcal{K}_m, t \in \mathcal{T}, s \tag{2k}$$

$$z_s \in \mathbb{R} \quad \forall s \in \mathcal{S} \tag{2l}$$

The objective function (2a) is the expected or "sample average" profit over a finite set of scenarios. The profit z_s in scenario s in equation (2b) includes two terms: the two positive terms account for revenues from the long-term and spot markets, while the two negative terms denote the cost to produce and transport volumes to markets. As a reminder, the only uncertain parameter is the spot price P_{mkts}, rendering Formulation (2) a stochastic MILP with objective function uncertainty only. Constraints (2c) ensure that the amount of production allocated to long-term wholesale market contracts (x^{Term}_{mcts}) adheres to the terms of the contract. Supply-demand balance constraints (2d) ensure that the total production equals the total quantities supplied to the markets. Constraints (2e) ensure that total supply equals the total amount transported to all market locations. Constraints (2f) govern supply limits by ensuring that total production is within lower and upper bounds. Side constraints (2g) capture potential mixed-integer requirements through the set $\mathcal{X}$. The remaining constraints are variable bounds.

To handle "separable" cross-market price elasticities, i.e., the situation when the volume supplied to market m_1 impacts/depresses the price in market $m_2(\neq m_1)$ as well, one could replace the term $\sum_{k\in\mathcal{K}_m} P_{mkts} y^{\text{Spot}}_{mkts}$ in (2b) with a more complex multivariate "staircase" representation $\sum_{k\in\mathcal{K}_m} P_{mkts} y^{\text{Spot}}_{mkts} - \sum_{m'\neq m}\sum_{k'\in\mathcal{K}_{m'}} \delta_{mm'k'ts} y^{\text{Spot}}_{m'k'ts}$. Here, $\delta_{mm'kts}$ denotes the per-unit price reduction in market m due to volumes sold in market m'. More sophisticated piecewise linear representations could be used to capture and linearize other nonlinear cross-market price relationships. For ease of exposition and due to the additional complexities associated with forecasting these nonlinear relationships, we will henceforth omit this interesting generalization.

Conditional value-at-risk formulation

Transitioning away from a purely risk-neutral mindset, a risk-averse decision maker could consider a two-stage risk-averse stochastic program that maximizes a convex combination of the expected profit and the expected "tail" profit or conditional value-at-risk ($\mathbb{CVAR}$) profit (Rockafellar, 2007). Conditional value-at-risk has emerged as an extremely popular risk metric due to its coherency and tractability. As such, it has been widely used in applications of optimization under uncertainty.

To arrive at a risk-averse formulation, let $\lambda\in[0,1]$ be a user-defined parameter governing the weight given to the expected value. When $\lambda=1$, the model reverts to the risk-neutral formulation (2), while $\lambda=0$ implies that the decision maker is only concerned with the expected "tail" profit. Let $\alpha\in(0,1)$ denote the risk-aversion parameter (or the α-quantile) in the $\mathbb{CVAR}$ calculation. In our maximization setting, $\mathbb{CVAR}_\alpha(X)$ denotes the expectation of a random variable (profit) X in the conditional distribution of its α-lower tail, e.g., the average of the lowest $\alpha=10\%$ profits. These assumptions lead to the following risk-averse MILP formulation:

$$\max_{\substack{\mathbf{x}^{\min},\nu^{\text{VaR}},\\ \boldsymbol{\ell},\mathbf{u},\mathbf{x},\mathbf{y},\mathbf{z}}} \lambda\sum_{s\in\mathcal{S}}\pi_s z_s + (1-\lambda)\left[\nu^{\text{VaR}} - \tfrac{1}{1-\alpha}\sum_{s\in\mathcal{S}}\pi_s\ell_s\right] \tag{3a}$$

$$\text{s.t.}\quad \ell_s \geq \nu^{\text{VaR}} - z_s, \ell_s\geq 0 \qquad \forall s\in\mathcal{S} \tag{3b}$$

$$\nu^{\text{VaR}}\in\mathbb{R} \tag{3c}$$

$$(2\text{b})-(2\text{l}) \tag{3d}$$

The additional decision variables z_s and ℓ_s denote the profit and the nonnegative tail loss in scenario s, while ν^{VaR} captures the value-at-risk at the α confidence level.

Distributionally robust model over a Wasserstein ball

Although Formulation (3) allows for some degree of risk-aversion, it is not "ambiguity-averse." That is, Formulation (3) assumes that an empirical probability distribution is known with certainty and then attempts to maximize the expected "tail" profit (and possibly other terms) with respect to this known distribution. In contrast, a distributionally robust model assumes that the distribution itself is unknown and belongs to a known family of distributions. After "centering" the family around the empirical data, a DRO approach over a Wasserstein ball maximizes the worst-case expected profit over this chosen family.

To arrive at a tractable DRO formulation over a Wasserstein ball, we describe the price vector $\mathbf{p}$ as a linear function $\mathbf{p}=\mathbf{Q}\boldsymbol{\xi}+\mathbf{q}$, where $\mathbf{Q}$ and $\mathbf{q}$ are a predefined matrix and vector that map an underlying basis vector $\boldsymbol{\xi}$ of uncertain parameters to $\mathbf{p}$. Furthermore, we must specify the Wasserstein radius $\varepsilon\in\mathbb{R}_+$, which defines the permissible deviation from the empirical data. Following (Xie, 2020, Prop. 2), one can formulate a two-stage DRO model as follows:

$$\max_{\substack{\mathbf{x}^{\min},\\ \mathbf{u},\mathbf{x},\mathbf{y},\mathbf{z}}} \sum_{s\in\mathcal{S}}\pi_s z_s - \varepsilon\sum_{s\in\mathcal{S}}\pi_s||\mathbf{Q}^\top\mathbf{y}^{\text{Spot}}_s||_{p^*} \tag{4a}$$

$$\text{s.t.}\quad (2\text{b})-(2\text{l}) \tag{4b}$$

Analogous to the $\mathbb{CVAR}$ term $(1-\lambda)\left[\nu^{\text{VaR}} - \frac{1}{1-\alpha}\sum_{s\in\mathcal{S}}\pi_s\ell_s\right]$ in Formulation (3), the term $\varepsilon\sum_{s\in\mathcal{S}}\pi_s||\mathbf{Q}^\top\mathbf{y}_s||_{p^*}$ in (4a) acts as a penalty on the spot allocation. The larger the ε radius, the larger the penalty on downside volatility. Setting $p=1$ or $p=\infty$, the dual norm $p^*=\infty$ or 1, respectively, and can therefore be represented using linear constraints.

Numerical Results

This section documents numerical results for a case study in the PJM electricity market. For simplicity, we assume a single market location $|\mathcal{M}|=1$, a planning horizon of $|\mathcal{T}|=365$ days, and $|\mathcal{S}|=100$ scenarios are available. Figure 1 depicts a truncated histogram of historical real-time hourly location marginal prices (LMPs) at PJM pricing node 48612 from 1 Jan 2021 through 1 May 2022 available at `http://dataminer2.pjm.com/feed/rt_hrl_lmps`. The histogram is truncated as prices spiked higher than \$200 in several hours. There is a single supply cost and no transportation cost.

We assume that $X^+_{mc}=0$ for all c (recall $|\mathcal{M}|=1$), the minimum and maximum supply satisfy $L_t=U_t=500$ MW for all t. For some perspective, the average capacity of a natural gas-fired combined-cycle power block is roughly 500 MW. There are no mixed-integer constraints in $\mathcal{X}$, i.e., all of our instances are LPs. The long-term contract demand curve has individual contracts up to 20 MW each, starting at a price of 38 \$/MWh, decreasing by 1 \$/MW with each subsequent contract, i.e., $W_{mc}=38-(c-1)$ and $X^{\max}_{mc}=20$ MW for all $c=1,\ldots,20$. The spot price elasticity curve has steps of width 25 MW, while the spot price decreases by 0.2 \$/MWh relative to the nominal parameter value in each step. That is, $P_{kts}=P_{1ts}-(k-1)0.2$ \$/MWh and $Y^{\text{Spot}}_{kts}=25$ MW for $k=1,\ldots,|K|=20$. We set $\lambda=0.01$, thus weighting the $\mathbb{CVAR}$ component of the objective function much more heavily than the expected profit.

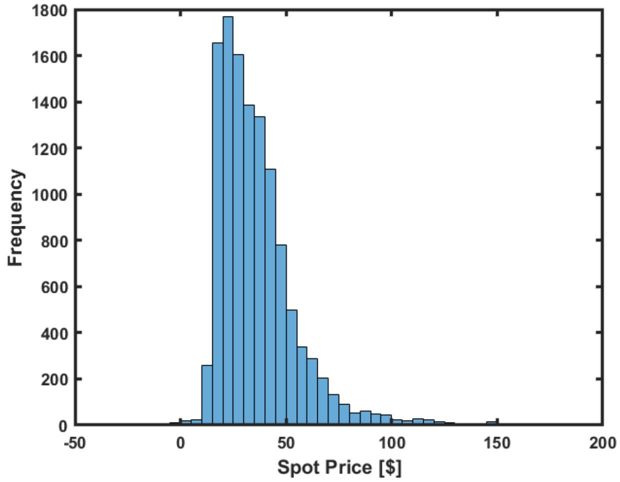

Figure 1: *Histogram of historical real-time hourly location marginal prices (LMP) at PJM node 48612 from 1 Jan 2021 through 1 May 2022.*

To compare the three formulations put forth in the previous section, we need to define several metrics:

- $\zeta(\mathbf{y}^{\text{Spot}})$ = Expected risk-neutral profit given a spot allocation decision $\mathbf{y}^{\text{Spot}}$
- $\chi_\alpha(\mathbf{y}^{\text{Spot}})$ = Expected α-tail profit ($\mathbb{CVAR}_\alpha$) given a spot allocation decision $\mathbf{y}^{\text{Spot}}$ and $\alpha \in (0,1)$
- ζ^{riskfree} = Expected risk-free profit. Note that $\zeta^{\text{riskfree}} = \zeta(\mathbf{0}) = \chi_\alpha(\mathbf{0}) \quad \forall \alpha \in (0,1)$
- $\Delta\zeta(\mathbf{y}^{\text{Spot}}) = \zeta(\mathbf{y}^{\text{Spot}}) - \zeta^{\text{riskfree}}$ = Expected increase in the risk-neutral profit relative to the risk-free profit given a spot allocation decision $\mathbf{y}^{\text{Spot}}$
- $\Delta\chi_\alpha(\mathbf{y}^{\text{Spot}}) = |\chi_\alpha(\mathbf{y}^{\text{Spot}}) - \zeta^{\text{riskfree}}|$ = Absolute value of the expected decrease in the α-tail profit relative to the risk-free profit given a spot allocation decision $\mathbf{y}^{\text{Spot}}$
- $\Delta\zeta(\mathbf{y}^{\text{Spot}})/\Delta\chi_\alpha(\mathbf{y}^{\text{Spot}})$ = Change in expected profit per change in risk (expected α-tail profit) as a function of the spot allocation $\mathbf{y}^{\text{Spot}}$

With these definitions, we can analyze the risk-reward tradeoff associated with allocating more supply to the spot market. Figure 2 depicts the "change in reward per change in risk" metric $\Delta\zeta(\mathbf{y}^{\text{Spot}})/\Delta\chi_\alpha(\mathbf{y}^{\text{Spot}})$ as a function of the spot allocation. The risk term in the denominator is measured using both the $\mathbb{CVAR}_{\alpha=0.05}$ and $\mathbb{CVAR}_{\alpha=0.10}$ metrics, which is also reflected in the legend labels. Intuitively, as one takes on more risk, the profit per unit risk decreases. When virtually all supply is allocated to essentially risk-free long-term contracts (see the left side of Figure 2), re-allocating a small amount of supply to the spot market results in a small decrease in the $\mathbb{CVAR}$ and a large increase in expected profit. More colloquially, a practitioner would say that taking on a small amount of risk (measured in terms of $\mathbb{CVAR}$) results in a large expected reward. Note that the horizontal axis in Figure 2 begins with a spot allocation of 20% because, at a 0% spot allocation, the "change in reward per change in risk" metric $\Delta\zeta(\mathbf{y}^{\text{Spot}})/\Delta\chi_\alpha(\mathbf{y}^{\text{Spot}})$ is extremely large due to a small denominator. As more supply is re-allocated to the spot market, however, this profit-per-unit-risk metric tapers off revealing that an additional unit of expected profit is only available by taking on a larger amount of risk.

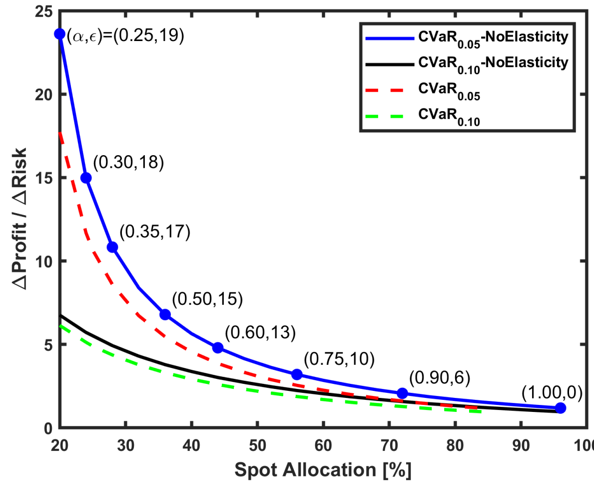

Figure 2: *Profit vs risk tradeoff curves. ΔProfit and ΔRisk are $\{$the expected profit increase$\}$ and $\{$the absolute value of the $\mathbb{CVAR}_\alpha$ decrease$\}$, respectively, relative to the risk-free allocation of committing all supply to long-term contracts. The (α,ε) values shown next to a subset of points refer to the $\mathbb{CVAR}_\alpha$ confidence level and the Wasserstein radius ε that induce the spot allocation on the horizontal axis. Two α values (0.05 and 0.10) are used in the ΔRisk calculation, while α values from 0 to 1 are tested to generate the four curves.*

The impact of including spot price elasticity is also illustrated. As stated in the previous section, the presence of spot price elasticity implies that, as the decision maker injects more supply into the market, prices may decrease below their nominal scenario value. This price (and hence profit) reduction relative to the "NoElasticity" curve is most noticeable on the left of Figure 2. This behavior occurs because, when only a small amount of supply is reserved for the spot market, it is able to capitalize on "high price events," which become less profitable as additional supply depresses their prices. Consequently, the numerator ΔProfit is lower relative to the "NoElasticity" assumption and the "change in reward per change in risk" metric decreases.

We now contrast the formulations. To this end, it is important to note that different values of the confidence value α and Wasserstein radius ε induce different long-term vs. spot market allocation decisions. These (α,ε) values are shown above a select number of points in Figure 2. As shown by the rightmost point in Figure 2, the risk-neutral formulation (2) is equivalent to setting the confidence level $\alpha = 1$ (recall that, for a random variable X, $\mathbb{CVAR}_1[X] = \mathbb{E}[X]$ by definition) or the Wasserstein radius $\varepsilon = 0$. Furthermore, in this example, the risk-neutral approach allocates 96% of supply to the spot market, but commits 4% of supply to long-term contracts. The reason why 4% is retained is because the expected spot price without elasticity is roughly 37.42 \$/MWh, while the first long-term contract is available for 38 \$/MWh. In agreement with intuition, as α decreases and ε increases (i.e., as we move from right to left in Figure 2), risk aversion increases and less supply is allocated to the volatile spot market in favor of greater price stability via long-term contracts.

Finally, a word on how to interpret the Wasserstein radius ε is in order. While the interpretation of the confidence level α is well understood as the lower α-tail expectation (when maximizing) in the $\mathbb{CVAR}_\alpha$ calculation (see, e.g., Rockafellar (2007)), the Wasserstein radius is perhaps less well known. Ramdas et al. (2017) provide some intuition. For a one-dimensional random variable, the ∞-Wasserstein distance, which is used in our example, between two bounded probability distributions can be interpreted as the maximum distance between the quantile functions of the two distributions. Thus, when $\varepsilon = 1$, one can interpret the DRO formulation (4) as imposing a limit of $\varepsilon = 1$ on the maximum distance between the quantile function of the empirical spot price distribution and that of the worst-case spot price distribution.

Conclusions and future research directions

In this work, we have attempted to make a case for applying DRO to the rather generic problem of balancing supply allocation to long-term contracts where price stability reigns versus spot markets where volatility can lead to large gains ... and losses. Focusing on a Genco's medium-term planning problem, we presented risk-neutral, risk-averse ($\mathbb{CVAR}$), and ambiguity-averse (DRO) formulations to address it. These formulations also included price elasticity components atypical in the majority of price-taker models. We then demonstrated how a risk-averse and an ambiguity-averse approach converge to the same decisions depending on the parameter values chosen. As always, the final long-term vs. spot market allocation decision depends on the risk/ambiguity aversion of the decision maker.

As for future research directions, it would be interesting to consider simultaneous uncertainty in the objective function and the constraints. Additionally, one could explore price setter behavior or game-theoretic models in which suppliers could exert market power and could therefore act as Nash-Cournot players.

References

Chen, R. and Paschalidis, I. C. (2018). A robust learning approach for regression models based on distributionally robust optimization. *Journal of Machine Learning Research*, 19(13).

Conejo, A. J., Garcia-Bertrand, R., Carrion, M., Caballero, Á., and de Andres, A. (2008). Optimal involvement in futures markets of a power producer. *IEEE Transactions on Power Systems*, 23(2):703–711.

Delage, E., Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612.

Esfahani, P. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166.

Fanzeres, B., Street, A., and Barroso, L. A. (2014). Contracting strategies for renewable generators: A hybrid stochastic and robust optimization approach. *IEEE Transactions on Power Systems*, 30(4):1825–1837.

Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *arXiv:1604.02199*.

Gao, J., Ning, C., and You, F. (2019). Data-driven distributionally robust optimization of shale gas supply chains under uncertainty. *AIChE Journal*, 65(3):947–963.

Hanasusanto, G. A. and Kuhn, D. (2018). Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research*, 66(3):849–869.

Kaye, R., Outhred, H., and Bannister, C. (1990). Forward contracts for the operation of an electricity industry under spot pricing. *IEEE Transactions on Power Systems*, 5(1):46–52.

Kirschen, D. S. and Strbac, G. (2018). *Fundamentals of power system economics*. John Wiley & Sons.

Kwon, R. H., Scott Rogers, J., and Yau, S. (2006). Stochastic programming models for replication of electricity forward contracts for industry. *Naval Research Logistics (NRL)*, 53(7):713–726.

Liu, B. and Yuan, Z. (2021). Multistage distributionally robust design of a renewable source processing network under uncertainty. *Ind Eng Chem Res*, 60(21):7883–7903.

Lorca, Á. and Prina, J. (2014). Power portfolio optimization considering locational electricity prices and risk management. *Electric Power Systems Research*, 109:80–89.

Pineda, S. and Conejo, A. J. (2012). Managing the financial risks of electricity producers using options. *Energy economics*, 34(6):2216–2227.

Rahimian, H. and Mehrotra, S. (2019). Distributionally robust optimization: A review. *arXiv:1908.05659*.

Ramdas, A., Trillos, N. G., and Cuturi, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2):47.

Rockafellar, R. T. (2007). Coherent approaches to risk in optimization under uncertainty. *Tutorials in Operations Research*, 3:38–61.

Sen, S., Yu, L., and Genc, T. (2006). A stochastic programming approach to power portfolio optimization. *Operations Research*, 54(1):55–72.

Shang, C. and You, F. (2018). Distributionally robust optimization for planning and scheduling under uncertainty. *Computers & Chemical Engineering*, 110:53–68.

Street, A., Barroso, L. A., Granville, S., and Pereira, M. V. (2009). Bidding strategy under uncertainty for risk-averse generator companies in a long-term forward contract auction. In *2009 IEEE PES General Meeting*, 1–8.

Wang, Z., Glynn, P. W., and Ye, Y. (2016). Likelihood robust optimization for data-driven problems. *Computational Management Science*, 13(2):241–261.

Xie, W. (2020). Tractable reformulations of two-stage distributionally robust linear programs over the type-∞ Wasserstein ball. *Operations Research Letters*, 48(4):513–523.

Yau, S., Kwon, R. H., Rogers, J. S., and Wu, D. (2011). Financial and operational decisions in the electricity sector: Contract portfolio optimization with the conditional value-at-risk criterion. *International Journal of Production Economics*, 134(1):67–77.