Les Cahiers du GERAD ISSN: 0711–2440

A distributionally robust optimization strategy for virtual bidding in two-settlement electricity markets

X. Audet, K. Qako, A. Lesage-Landry

G–2025–31

April 2025

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2025
– Bibliothèque et Archives Canada, 2025

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2025
– Library and Archives Canada, 2025

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

A distributionally robust optimization strategy for virtual bidding in two-settlement electricity markets

Xavier Audet [a, b, c, d]

Kliti Qako [d]

Antoine Lesage-Landry [a, b, c]

[a] *GERAD, Montréal (Qc), Canada, H3T 1J4*

[b] *Polytechnique Montréal, Montréal (Qc), Canada, H3T 0A3*

[c] *Mila – Institut québécois de l'intelligence artificielle, Montréal (Qc), Canada, H2S 3H1*

[d] *Hydro-Québec, Montréal (Qc), Canada, H2Z 1A4*

xavier.audet@polymtl.ca
qako.kliti@hydroquebec.com
antoine.lesage-landry@polymtl.ca

April 2025
Les Cahiers du GERAD
G–2025–31

Abstract : This work presents a new strategy to virtual bidding based on distributionally robust optimization (DRO) using a Wasserstein distance. Virtual bidding, a mechanism used in two-settlement electricity markets, allows participants to arbitrage price differences between day-ahead and real-time markets. Traditional optimization methods for virtual bidding often rely on precise probabilistic models of market behaviour, which are unavailable in practice due to the inherent complexity and volatility of electricity markets. To tackle these challenges, this work formulates the virtual bidding problem as a DRO problem, incorporating conditional value at risk (CVaR) into the objective to manage downside risk under volatile conditions. Tractable reformulations which can be efficiently solved to optimality are provided. The proposed strategy is developed and tuned using a 12-month training set to identify optimal parameters. The strategy is evaluated on historical pricing data from the New York Independent System Operator (NYISO) on an 8-month testing set. The results show improved performance over benchmarks, achieving higher Sharpe and Calmar ratios, as well as increased profit per MWh. Through this DRO framework, a more reliable virtual bidding strategy that enhances profitability while effectively managing risk in uncertain market environments is presented.

Keywords: Virtual Bidding, distributionally robust optimization, stochastic optimization, wholesale electricity market, two-settlement electricity market

Acknowledgements: We would like to thank Hydro-Québec for providing us with the computing resources of Lab CASTOR for this project (128 Intel(R) Xeon(R) Platinum 8375C CPUs @ 2.90GHz with 4.2 TB of RAM over a period of 7 days 24/7).

This work was funded by Mitacs, Hydro-Québec, and theNatural Sciences and Engineering Research Council of Canada (NSERC).

1 Introduction

Virtual bidding is a financial mechanism used in two-settlement electricity markets that allows participants to arbitrage price differences between the day-ahead and real-time markets. This practice aims to enhance market efficiency by inducing price convergence between day-ahead and real-time market prices. This convergence encourages better resource allocation, as prices align closer to real-time generation and demand costs [16]. This can in turn enhance grid efficiency by ensuring that generators are dispatched according to actual needs, thus supporting the overall reliability of the electric power system [13].

The effectiveness of virtual bidding strategies heavily depends on the availability of accurate models of market behaviour. These are often challenging to obtain due to the complex nature, the inherent uncertainty, and the volatility of electricity markets.

Despite its potential benefits, virtual bidding remains a challenging practice due to the high level of uncertainty in electricity markets and the volatility of the prices. Price discrepancies between the day-ahead and real-time markets are influenced by numerous factors, including load forecasting errors, generator outages, fuel price fluctuations, and strategic bidding by market participants. These can lead to significant financial risks, making it crucial to develop robust strategies to improve bidding performance while mitigating the adverse effect of volatility. This paper addresses this gap by proposing a new approach that integrates distributionally robust optimization for virtual bidding under uncertainty.

This work focuses on electricity markets that support virtual bidding, primarily concentrating on independent system operators (ISOs) in the United States, such as the New York Independent System Operator (NYISO). These markets allow for arbitrage opportunities between the day-ahead and real-time markets. In contrast, most European markets, including the OMI Group, responsible for the Iberian Peninsula [27], and the National Energy System Operator (NESO) in the United Kingdom [10], mainly operate through the day-ahead and intra-day markets, with no equivalent real-time market like in the United States. While intra-day markets in Europe allow some arbitrage opportunities, they do not directly allow for virtual bidding strategies like the DART spread does due to the absence of a real-time market for price adjustments. As such, virtual bidding, as formulated in this work, is most applicable to markets with both day-ahead and real-time market structures, such as those found in the United States ISOs.

1.1 Related work

Various approaches have been proposed to model uncertainty in electricity markets. Distributionally robust optimization (DRO) trading strategies for renewable energy producers are considered in [29]. The author highlights the importance of robust optimization to deal with the uncertainty inherent to renewable sources of energy. The authors of [11] focus on forecasting electricity DART spikes, providing valuable insight into the prediction of extreme events in electricity markets. Their methodology could be, for example, integrated into a DRO framework to enhance decision-making under uncertainty. A two-stage stochastic bidding strategy for physical market participants with virtual bidding capacities in day-ahead electricity markets is proposed in [21]. Their approach aims to balance the physical generation profit and the virtual transaction profit. Virtual bidding can improve market efficiency, but there is also a risk associated to it, as shown in [4]. The authors discuss the implications of virtual bidding in electricity markets, highlighting both the benefits and potential pitfalls. Their analysis provides a balanced view of virtual bidding strategies. A data-driven convergence bidding strategy based on reverse engineering of market participants' performance is proposed in [32]. A case study on the California ISO demonstrating the practical applicability of the approach is also presented. Algorithmic bidding strategies based on historical data are proposed and evaluated using the cumulative profit and the Sharpe ratio in [2]. The authors show that their strategy outperforms standard benchmarks and the S&P 500 index over the same period. A significant body of literature also explores different uncertainty modelling approaches beyond traditional stochastic optimization.

Robust optimization methods, for instance, focuses on worst-case scenarios to ensure optimal decisions under the most adverse conditions [22]. More recently, DRO has gained traction as an alternative that provides a middle ground between stochastic and robust optimization by considering a family of probability distributions in a data-driven fashion instead of a single assumed or known one [17, 24]. In electricity markets, DRO has been applied to problems such as generation scheduling [29]. However, to the best of the authors' knowledge, its application to virtual bidding remains unexplored. Many virtual bidding strategies rely on optimization techniques and machine learning methods [2, 21, 32]. There is also a growing body of literature focused on DRO reformulations in various fields [5, 7, 24, 28, 29]. However, no existing strategy integrates all these elements simultaneously, leaving a gap in the literature for a virtual bidding approach that combines DRO optimization with machine learning-inspired parameter optimization to adequately model uncertainty and risk in electricity markets.

1.2 Contributions

This work proposes an uncertainty- and risk-aware virtual bidding methodology to account for two-settlement electricity market uncertainty and volatility. Our approach is data-driven to further account for the lack of generally available system model. Specifically, we formulate the virtual bidding problem as a DRO with a Wasserstein distance [24] and provide a tractable reformulation that can be efficiently solved. The strategy is obtained by identifying the optimal parameters of the model over a 12-month training set. The effectiveness of our approach is shown by using historical data from NYISO on a 8-month testing set, showing that DRO-based strategies outperform traditional methods, namely, the equally weighted selling portfolio and scenario-based stochastic optimization approaches especially in instances with high market volatility and uncertainty. The specific contributions of our work are as follows:

- We formulate a distributionally robust optimization model that explicitly accounts for market uncertainty and volatility.
- We propose an online bidding strategy with an integrated pipeline for efficiently tuning model parameters, making the strategy adaptable to changing market and network conditions.

To evaluate the effectiveness of our strategy, we conduct a numerical case study using 8 months of historical data from the NYISO. While similar studies using NYISO data exist, our case study provides new insights by comparing our bidding strategy against other uncertainty-aware models and existing benchmarks, highlighting the advantages in terms of both profitability and risk mitigation. This validation based on cumulative profits, profits per invested MWh, Sharpe ratio, and Calmar ratio, demonstrates the practical applicability and effectiveness of our proposed approach.

1.3 Organization

The remainder of this paper is organized as follows. Section 2 introduces virtual bidding in two-settlement wholesale electricity markets. Section 3 builds the model to solve the virtual bidding problem. Section 4 proposes a bidding strategy based on our model and includes tuning considerations. Section 5 provides numerical results and analyzes the performance of the strategy. Finally, Section 6 concludes the paper and outlines directions for future research.

2 Virtual bidding

Virtual bidding, also known as convergence bidding, allows market participants to submit financial bids without physically delivering or consuming electric energy. In a two-settlement electricity market, which consists of a day-ahead and a real-time market, virtual bids aim to capitalize on the price differences between these two markets [15]. Participants place bids in the day-ahead market as if they were buyers (loads) or sellers (generators) and then settle their positions in the real-time market. In doing so, virtual bidders help align the day-ahead prices with real-time ones, thereby improving market efficiency and reducing price volatility.

When submitting bids, participants must specify both a price and a quantity for a particular bus of the grid, e.g., in California Independent System Operator (CAISO) [3], Midcontinent Independent System Operator (MISO) [23], PJM Interconnection LLC (PJM) [30], ISO New England (ISO-NE) [14], and Southwest Power Pool (SPP) [33] markets, or a region of aggregated buses, e.g., in NYISO [26] and Electric Reliability Council of Texas (ERCOT) [8] depending on the grid operator, at a specific hour of the day. The bids are cleared if the buy price is greater than or equal to the price set by the grid operator in the day-ahead market or if the sell price is less than or equal to the day-ahead market price. To avoid the computational burden of mixed-integer programming, we adopt a price-taker approach, where we always submit a high price for buying and a low price for selling. This approach ensures that our bids clear every time and allows us to focus on optimizing the energy quantity rather than the bid price itself. Because our strategy does not seek to influence market prices but rather takes them as given, this price-taker assumption is reasonable and effective for simplifying the bidding process. Previous studies have also adopted a similar approach to reduce computational complexity in energy markets [32].

Let the power system be modelled as the graph $(\mathcal{N}, \mathcal{L})$ where $\mathcal{N} \subset \mathbb{N}$ denotes the set of buses and $\mathcal{L} \subset \mathcal{N} \times \mathcal{N}$ represents the set of transmission lines connecting these buses. The quantity vector $\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}$ collects the bid energy quantities (in MWh) submitted at all buses at a specific hour $t \in \mathcal{T} = \{0, 1, ..., 23\}$ of the considered day. To ensure operational feasibility, we impose a limit $L > 0$ on the total energy that can be bid across all buses in each hour, expressed as:

$$\|\mathbf{q}_t\|_1 \leq L, \qquad \forall t \in \mathcal{T}.$$

We suppose that the limit L ensures that our bids do not affect prices in the considered market. The impact of virtual bids on prices (price-makers) is a topic for future work. It is explored, for example, in [18].

Let $\mathcal{S} \subset \mathbb{R}^{|\mathcal{N}|}$ be the uncertainty set of the random variable $\mathbf{s}_t \in \mathcal{S}$ representing the DART spreads, for all buses at time t. We model the spread $\mathbf{s}_t$ as a random variable with distribution $\mathbb{P} \in \mathcal{P}$, where $\mathcal{P}$ is the set of probability distribution on $\mathcal{S}$. The virtual bidding problem takes the following form:

$$\min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}} \quad \mathbb{E}^{\mathbb{P}}\left[\sum_{t \in \mathcal{T}} -\mathbf{s}_t^\top \mathbf{q}_t\right] \tag{1a}$$

$$\text{s.t.} \quad \|\mathbf{q}_t\|_1 \leq L, \qquad \forall t \in \mathcal{T}, \tag{1b}$$

where the expectation is taken with respect to $\mathbf{s}_t$.

Given the complexity of the market, i.e., its dependency on many factors such as the weather, the availability of energy, and consumer behaviours, the distribution $\mathbb{P}$ is unknown. In this work, we are interested in a data-driven approach. Using a set $\mathcal{D}$ of historical DART spreads $\mathbf{s}_t^d$, the distribution $\mathbb{P}$ can be approximated with the discrete empirical probability distribution:

$$\hat{\mathbb{P}}_{\mathcal{D}} = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \delta_{\mathbf{s}_t^d},$$

where $\delta_{\mathbf{s}_t^d}$ is the Dirac measure giving a probability mass $\frac{1}{|\mathcal{D}|}$ to each observed data point $\mathbf{s}_t^d \in \mathcal{D}$. This expression reflects our estimation of the distribution based on the observed data. This enables us, for example, to approximate (1) as:

$$\min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}} \quad \mathbb{E}^{\hat{\mathbb{P}}_{\mathcal{D}}}\left[-\sum_{t \in \mathcal{T}} \mathbf{s}_t^\top \mathbf{q}_t\right] := -\frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \left[\sum_{t \in \mathcal{T}} {\mathbf{s}_t^d}^\top \mathbf{q}_t\right] \tag{2a}$$

$$\text{s.t.} \quad \|\mathbf{q}_t\|_1 \leq L, \qquad \forall t \in \mathcal{T}, \tag{2b}$$

which we later use as a benchmark in our numerical study. We refer to the scenario-based, single-stage optimization problem (2) as the stochastic optimization (`SO`) model.

3 Volatility- and uncertainty-aware bidding model

Volatility in electricity markets refers to the significant fluctuations in prices over time, which introduce risks for market participants [25]. Concurrently, variability in market conditions encompasses broader uncertainties, such as changes in generation, demand, or external factors affecting the distribution of prices [19]. In our approach, we address both challenges using a combination of the conditional value at risk (CVaR) and DRO, respectively. CVaR focuses on tail risks, providing a safeguard against extreme price movements which is a direct consequence of volatility. By minimizing potential losses in the worst-case scenarios, CVaR helps manage the impact of sharp price drops. In parallel, DRO considers a range of possible market conditions by defining an ambiguity set, which captures variability in the market's underlying distribution. This ensures that our virtual bidding strategy remains robust not only to known price risks but also to shifts and uncertainties in future market conditions.

3.1 Volatility mitigation

The conditional value at risk (CVaR), or expected shortfall, is a risk measure commonly used to evaluate investment strategies [34] that captures the tail risk beyond a given confidence level α. The CVaR is defined as the average of the worst-case outcomes beyond a specified confidence level [31], e.g., the average return of the worst 10%. To account for the volatility in our model, we integrate the CVaR in the objective function of the single-stage stochastic program (1). The resulting formulation minimizes a combination of the expected loss and the CVaR of the loss. Let $\alpha \in (0,1]$ be the confidence level used for the CVaR, and representing the portion of tail risk considered. Let $\rho \in [0,1]$ be the risk-aversion factor which we use to balance between minimizing the expected loss and the CVaR. The problem formulation becomes:

$$\begin{aligned} \min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}} \quad & \rho \mathbb{E}^{\mathbb{P}} \left[-\sum_{t \in \mathcal{T}} \mathbf{s}_t^\top \mathbf{q}_t \right] + (1-\rho) \mathbb{P}\text{-CVaR}_\alpha \left[-\sum_{t \in \mathcal{T}} \mathbf{s}_t^\top \mathbf{q}_t \right], && \text{(3a)} \\ \text{s.t.} \quad & \|\mathbf{q}_t\|_1 \le L, \qquad \forall t \in \mathcal{T}, && \text{(3b)} \end{aligned}$$

where $\mathbb{P}\text{-CVaR}_\alpha$ denotes the CVaR-operator for a probability distribution $\mathbb{P}$ and a risk level α. Following [31], CVaR can be defined as a minimization problem:

$$\min_{\tau \in \mathbb{R}} \tau + \mathbb{E}^{\mathbb{P}} \left[\frac{1}{\alpha} \max \left\{ -\sum_{t \in \mathcal{T}} \mathbf{s}_t^\top \mathbf{q}_t - \tau,\ 0 \right\} \right], \tag{4}$$

where $\tau \in \mathbb{R}$ represents the CVaR threshold, which acts as a quantile of the loss distribution. It marks the loss level beyond which the tail risk is considered. Substituting in the CVaR definition (4), we obtain:

$$\begin{aligned} \min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}} \quad & \rho \mathbb{E}^{\mathbb{P}} \left[-\sum_{t \in \mathcal{T}} \mathbf{s}_t^\top \mathbf{q}_t \right] + (1-\rho) \min_{\tau \in \mathbb{R}} \mathbb{E}^{\mathbb{P}} \left[\tau + \frac{1}{\alpha} \max \left\{ -\sum_{t \in \mathcal{T}} \mathbf{s}_t^\top \mathbf{q}_t - \tau,\ 0 \right\} \right] && \text{(5a)} \\ \text{s.t.} \quad & \|\mathbf{q}_t\|_1 \le L, \qquad \forall t \in \mathcal{T}. && \text{(5b)} \end{aligned}$$

Problem (5) can be re-written as the following linear program using the auxiliary variable y^d, which represents the excess loss above threshold τ in each scenario $d \in \mathcal{D}$:

$$\begin{aligned} \min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}, \tau \in \mathbb{R}, y^d \in \mathbb{R}} \quad & -\rho \left(\frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \sum_{t \in \mathcal{T}} {\mathbf{s}_t^d}^\top \mathbf{q}_t \right) \quad + (1-\rho) \left(\tau + \frac{1}{|\mathcal{D}|} \alpha \sum_{d \in \mathcal{D}} y^d \right) && \text{(6a)} \\ \text{s.t.} \quad & \|\mathbf{q}_t\|_1 \le L, \qquad \forall t \in \mathcal{T}, && \text{(6b)} \\ & y^d + \tau + \sum_{t \in \mathcal{T}} {\mathbf{s}_t^d}^\top \mathbf{q}_t \ge 0, \qquad \forall d \in \mathcal{D}, && \text{(6c)} \\ & y^d \ge 0, \qquad \forall d \in \mathcal{D}. && \text{(6d)} \end{aligned}$$

The resulting is a model that incorporates risk into the optimization process. We refer to (6) as the `SO-CVaR` model and we use it later to benchmark our approach.

3.2 Distributional uncertainty modelling

DRO provides a framework for decision-making under uncertainty by considering a set of possible probability distributions rather than relying on the empirical distribution like in [2]. This approach aims to ensure that solutions remain robust against variability in the underlying data. To define our ambiguity set, we use the Wasserstein distance $d_W(\mathbb{W}, \mathbb{V})$, which quantifies the dissimilarity between probability distributions $\mathbb{W} \in \mathcal{P}$ and $\mathbb{V} \in \mathcal{V}$. Specifically, let s_1 and s_2 be the marginals of $\mathbb{W}$ and $\mathbb{V}$, respectively, and $P(\mathcal{S} \times \mathcal{S})$ denote the set of all probability measures on $\mathcal{S} \times \mathcal{S}$. The order-$z$ Wasserstein distance [28] with respect to some norm $\|\cdot\|$ is :

$$d_W(\mathbb{W}, \mathbb{V}) = \left(\min_{\pi \in P(\mathcal{S} \times \mathcal{S})} \int_{\mathcal{S} \times \mathcal{S}} \|s_1 - s_2\|^z \, \mathrm{d}\pi(s_1, s_2) \right)^{1/z}.$$

By constructing a Wasserstein ball of radius $\epsilon > 0$ centred around the empirical distribution:

$$\mathcal{B}_\epsilon(\hat{\mathbb{P}}_\mathcal{D}) := \left\{ \mathbb{W} \in \mathcal{P} \mid d_W(\mathbb{W}, \hat{\mathbb{P}}_\mathcal{D}) \le \epsilon \right\},$$

we can delineate a set of plausible distributions that captures the uncertainty inherent to the data. In this context, we formulate a DRO model to maximize the virtual bidding benefits. The distributionally robust reformulation of (1) is given by:

$$\min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}} \max_{\mathbb{W} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_\mathcal{D})} \quad \mathbb{E}^{\mathbb{W}} \left[-\sum_{t \in \mathcal{T}} \mathbf{s}_t^\top \mathbf{q}_t \right] \tag{7a}$$

$$\text{s.t.} \quad \|\mathbf{q}_t\|_1 \le L, \qquad \forall t \in \mathcal{T}. \tag{7b}$$

Before we proceed to the reformulation of the inner maximization of (7), we define, for some vector $\mathbf{x} \in \mathbb{R}^{|\mathcal{N}|}$, the dual norm $\|\cdot\|_*$ of the norm $\|\cdot\|$ as:

$$\|\mathbf{x}\|_* = \sup_{\|\mathbf{y}\| \le 1} \mathbf{x}^\top \mathbf{y}.$$

In our context, we choose the Euclidean norm in the Wasserstein distance. For the Euclidean norm, the dual norm is equal to the norm itself, i.e., $\|\mathbf{q}_t\|_* = \|\mathbf{q}_t\|_2$. This is done because it promotes a strategy that distributes bids across the different buses on the grid. We also choose the order $z = 1$ that corresponds to the Wasserstein-1 distance, also known as the earth mover distance (EMD). It is computationally simpler compared to higher-order Wasserstein distances.

Because the objective function of (7) is linear, we can re-express the inner maximization as [24, Remark 6.6]:

$$\min_{\lambda \in \mathbb{R}} \quad -\frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \sum_{t \in \mathcal{T}} {\mathbf{s}_t^d}^\top \mathbf{q}_t + \epsilon \lambda, \tag{8a}$$

$$\text{s.t.} \quad \|\mathbf{q}_t\|_* \le \lambda, \qquad \forall t \in \mathcal{T}. \tag{8b}$$

Substituting (8b) in (7) leads to:

$$\min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}, \lambda \in \mathbb{R}} \quad -\frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \sum_{t \in \mathcal{T}} {\mathbf{s}_t^d}^\top \mathbf{q}_t + \epsilon \lambda, \tag{8c}$$

$$\text{s.t.} \quad \|\mathbf{q}_t\|_2 \le \lambda, \qquad \forall t \in \mathcal{T}, \tag{8d}$$

$$\|\mathbf{q}_t\|_1 \le L, \qquad \forall t \in \mathcal{T}. \tag{8e}$$

Problem (8e) provides a distributionally robust virtual bidding model that is readily solvable with off-the-shelf solvers. For a linear problem like ours, the DRO reformulation is equivalent to the original problem with an added regulizer [5]. We refer to (8e) as the `DRO` model and we use it as well to benchmark our approach.

3.3 Integrated volatility-and uncertainty-aware virtual bidding

We now present our integrated bidding model. We combine the CVaR to account for price volatility and DRO to handle uncertainty in the market's underlying distribution in a data-driven way. CVaR protects against extreme price changes, while DRO ensures robustness to shifts in market conditions. This combined model addresses both difficulties in a single problem. Following the approach shown in [24], we can write the risk-averse problem (5) as:

$$\min_{\mathbf{q}_t \in \mathbb{R}^{|\mathcal{N}|}, \tau \in \mathbb{R}} \quad \mathbb{E}^{\mathbb{P}}\Big[\max_{k \in \mathcal{K}} a_k \left(\sum_{t \in \mathcal{T}} -\mathbf{s}_t^\top \mathbf{q}_t\right) + b_k \tau\Big], \tag{9a}$$

$$\text{s.t.} \quad \|\mathbf{q}_t\|_1 \leq L, \qquad \forall t \in \mathcal{T}, \tag{9b}$$

where $\mathcal{K} = \{1, 2\}$, $a_1 = -\rho$, $a_2 = -\rho - \frac{1-\rho}{\alpha}$, $b_1 = 1 - \rho$, $b_2 = (1 - \rho)(1 - \frac{1}{\alpha})$. Although the exact distribution $\mathbb{P}$ remains unknown, we propose using box constraints following [7] to define the uncertainty set $\mathcal{S}$ as the range within which the uncertain DART spread vector $\mathbf{s}_t$ is allowed to vary. We set:

$$\mathcal{S} = \{\mathbf{s}_t \in \mathbb{R}^{|\mathcal{N}|} : |s_{i,t}| \leq \Lambda, \quad \forall i \in \mathcal{N}, t \in \mathcal{T}\}, \tag{10}$$

where each component of $\mathbf{s}_t$, $s_{i,t}$, is bounded by the maximum value $\Lambda > 0$, ensuring that all elements of $\mathbf{s}_t$ remain within this fixed range. The distributionally robust counterpart of (9) with respect to the Wasserstein ambiguity set $\mathcal{B}_\epsilon(\hat{\mathbb{P}}_{\mathcal{D}})$ is given by:

$$\min_{\mathbf{q}_t, \tau} \max_{\mathbb{W} \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_{\mathcal{D}})} \mathbb{E}^{\mathbb{W}}\Big[\max_{k \in \mathcal{K}} a_k \left(\sum_{t \in \mathcal{T}} -\mathbf{s}_t^\top \mathbf{q}_t\right) + b_k \tau\Big] \tag{11a}$$

$$\text{s.t.} \quad \|\mathbf{q}_t\|_1 \leq L, \qquad \forall t \in \mathcal{T}. \tag{11b}$$

Using [24, (27)] and setting the vector $\boldsymbol{\nu}_k^d \in \mathbb{R}^{|\mathcal{N}|}$ and the scalar $\mu_k^d \in \mathbb{R}$ as variables associated to the uncertainty set (10) as in [7] allows for the readily implementable form:

$$\min_{\mathbf{q}_t, \tau, \lambda, x^d, \boldsymbol{\nu}_k^d, \mu_k^d} \quad \lambda \epsilon + \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} x^d, \tag{12a}$$

$$\text{s.t.} \quad b_k \tau + a_k \left(\sum_{t \in \mathcal{T}} {\mathbf{s}_t^d}^\top \mathbf{q}_t\right) + {\boldsymbol{\nu}_k^d}^\top \mathbf{s}_t^d + \Lambda \mu_k^d \leq x^d, \qquad \forall d \in \mathcal{D}, k \in \mathcal{K}, \tag{12b}$$

$$\|\boldsymbol{\nu}_k^d - a_k \mathbf{q}_t\|_2 \leq \lambda, \qquad \forall d \in \mathcal{D}, k \in \mathcal{K}, t \in \mathcal{T}, \tag{12c}$$

$$\mu_k^d \geq \|\boldsymbol{\nu}_k^d\|_1, \qquad d \in \mathcal{D}, k \in \mathcal{K}, \tag{12d}$$

$$\|\mathbf{q}_t\|_1 \leq L, \qquad \forall t \in \mathcal{T}. \tag{12e}$$

The detailed derivation to obtain (12) given the uncertainty set (10) is provided in the Appendix. This formulation provides a distributionally robust optimization framework that balances profit maximization and risk mitigation in addition to accounting for the uncertainty. The choice of the Euclidean norm in (12) aligns with the approach used to obtain (8e). We refer to (12) as the `DRO-CVaR` model.

4 Online virtual bidding strategy

We devise a rolling virtual bidding strategy centred on our model (12). Our strategy is based on a data set $\mathcal{D}$ updated daily to determine bids for the next day. To implement our model within a bidding strategy, we begin by selecting the most relevant data for each day that will define the dataset $\mathcal{D}$. Next, we optimize the hyperparameters over a training period to ensure the best model performance.

4.1 Selection of similar days

A key challenge in data-driven optimization is determining which historical data will represent the future best. In our study, we carefully select relevant historical data by identifying dates with load and generation profiles similar to the target bidding day. For each target date, we consider a two-year rolling window of past data and calculate the similarity between the target bidding date and each day within the historical window.

Let the similarity metric $\Gamma : (\mathbb{R}^{24} \times \mathbb{R}) \times (\mathbb{R}^{24} \times \mathbb{R}) \to \mathbb{R}$ measure the similarity between two days. The load profile $\mathbf{p} \in \mathbb{R}^{24}$ collects the total network load (across all buses) for each timestep over the day in a vector, while the offline generation capacity $\theta \in \mathbb{R}$ reflects the total megawatts (MW) of generating capacity forecasted to be offline for the ISO throughout the day. The similarity metric combines the Euclidean distance between load profiles and the absolute difference in offline capacities in addition to a penalty $\zeta > 0$ for splitting weekdays with weekends. The Γ metric is defined as:

$$\Gamma\left((\mathbf{p}_{\text{target}}, \theta_{\text{target}}), (\mathbf{p}, \theta)\right) = 2\|\mathbf{p}_{\text{target}} - \mathbf{p}\|_2 + |\theta_{\text{target}} - \theta| + \zeta,$$

where $\zeta = 0$ if both days are either weekdays or weekends, and $\zeta = 1000$ otherwise. The fixed values of ζ ensure the penalty significantly influences Γ without excluding the possibility that a weekend day is more similar to a target day than any weekday, and vice versa. The Γ metric ranks past days by similarity, with smaller values indicating more similar days. Finally, we remark that other comparison methods could be used in our approach.

4.2 Strategy execution

We use Γ to build the relevant historical data set $\mathcal{D}_j$ to be used by our model at day j. We discuss the number of data to include in $\mathcal{D}_j$ in Section 4.3. We then solve (12) using $\mathcal{D}_j$ and commit the resulting bids for day j. For example, consider bids to be placed for $j =$ May 13$^{\text{th}}$ on May 12$^{\text{th}}$. Let us assume $|\mathcal{D}| = 3$, evaluating Γ on the month of data preceding j, we determine that the most similar days to May 13$^{\text{th}}$ are April 30$^{\text{th}}$, May 3$^{\text{rd}}$, and May 7$^{\text{th}}$ and we form $\mathcal{D}_j$ accordingly. The process is visualized in Figure 1.

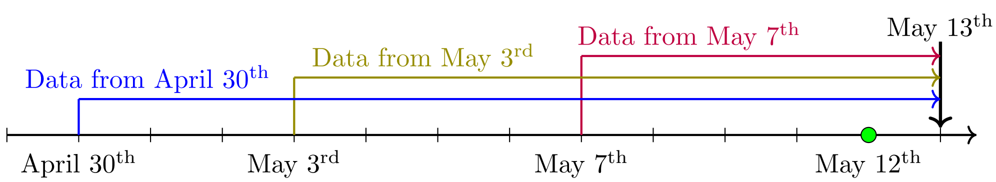


Figure 1: Strategy execution example for bids to be placed on May 12^{th} for the next day

Finally, we repeat this process for all bidding days $j \in \{1, 2, ..., J\}$, adding each day one more data point in the pipeline and removing one. The proposed strategy is deployed as an online process, updating daily both the dataset $\mathcal{D}$ and the broader dataset from which $\mathcal{D}$ is constructed. This broader dataset consists of a rolling two-year window that is updated each day to incorporate the most recent data while discarding the oldest observations. Within this window, $\mathcal{D}$ is dynamically selected based on the similarity of past days to the current one. This ensures that the model continuously adapts to the most relevant information while accounting for real-world changes such as evolving demand, climate variability, and grid modifications.

4.3 Strategy tuning

In this section, we describe our bidding strategy and tuning method, which we also apply to the benchmark models for a fair comparison. The hyperparameters of each model are optimized on a training set comprising 12 months of historical DART spreads from NYISO from February 1st, 2023, to January 31st, 2024 with the rolling window going back as far as February 1st, 2021 when tuning starts. The online bidding strategy is then evaluated on a separate 8-month testing set illustrated in Figure 2.

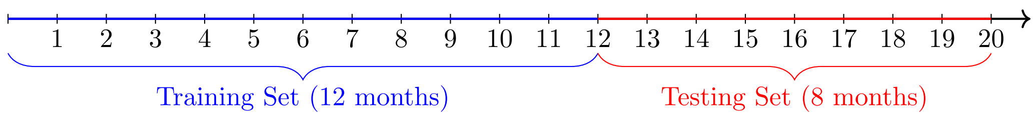


Figure 2: Training and testing sets

Problem (12) includes hyperparameters ϵ, ρ, $|\mathcal{D}|$, and Λ. Tuning is carried out using Optuna [1] over the training set. Its objective is to maximize the annualized Calmar ratio C. The Calmar ratio measures the trade-off between risk and return by comparing the annualized return of a bidding strategy to its maximum drawdown and provides an indication of the risk-adjusted performance of the strategy [20]. We first initialize a portfolio that has a value of \$1M, denoted as $v_0 = 10^6$.

The daily returns r_j for each day $j \in \{1, 2, ..., J\}$ are calculated as the total daily profit or loss from all bids, given by:

$$r_j = \sum_{t \in \mathcal{T}} -\mathbf{s}_t^\top \mathbf{q}_t.$$

The returns are then scaled relative to the value of the portfolio from the previous day v_{j-1} to obtain the scaled return η_j, capturing the proportionate growth or decline in the portfolio:

$$\eta_j = \frac{r_j}{v_{j-1}}. \tag{13}$$

The portfolio value is updated iteratively by adding the daily return:

$$v_j = v_{j-1} + r_j.$$

The annualized return $R_{\text{annualized}}$ is then calculated as the cumulative return for the year, scaled to a full 365-day period:

$$R_{\text{annualized}} = \left(\prod_{j=1}^{J} (1 + \eta_j) \right)^{\frac{365}{J}} - 1, \tag{14}$$

where J is the total number of trading days. The maximum drawdown (MDD) is defined as the largest peak-to-trough decline during the trading period:

$$\text{MDD} = \max \left(\frac{v_{\text{peak}} - v_{\text{trough}}}{v_{\text{peak}}} \right), \tag{15}$$

where v_{peak} is the highest portfolio value before a drawdown, and v_{trough} is the lowest value reached during the drawdown. Finally, combining (14) and (15), the Calmar ratio C is computed as:

$$C = \frac{R_{\text{annualized}}}{\text{MDD}}. \tag{16}$$

In our numerical example, these calculations are performed using the `empyri- cal` library [9]. The use of the Calmar ratio in our analysis allows for a robust evaluation of the portfolio's performance in high-risk environments, such as those observed in electricity markets.

The Calmar ratio is chosen to tune our model because it is a recognized tool for measuring risk-adjusted performance of investment strategies, particularly by focusing on downside risk [6]. This makes it a good objective to select hyperparameters during tuning. Each model is tuned on the same

Table 1: Hyperparameter range for Optuna on the 12-month training set

Hyperparameter	**min**	**max**
$\lvert\mathcal{D}\rvert$	2	100
ϵ	5	50
ρ	0.2	0.8
Λ	2000	5000

training set for 1000 trials with the quantity limit constraint set to $L = 400$ MWh. The tuning is parallelized and performed on 128 Intel(R) Xeon(R) Platinum 8375C CPUs @ 2.90GHz with 4.2 TB of RAM. We let Optuna try values for the hyperparameters in the ranges provided in Table 1 with the objective of maximizing the Calmar ratio over the 12-month testing set.

For `SO-CVaR` and `DRO-CVaR`, the hyperparameter α is fixed to 10% to represent a predefined level of risk aversion, commonly used to focus on the worst 10% of outcomes. This approach ensures interpretability and stability, as optimizing α could lead to overly conservative or risk-seeking behaviours. Additionally, fixing α simplifies the hyperparameter optimization process, reducing its dimensionality and leveraging contextual information on acceptable risk levels in the electricity trading sector.

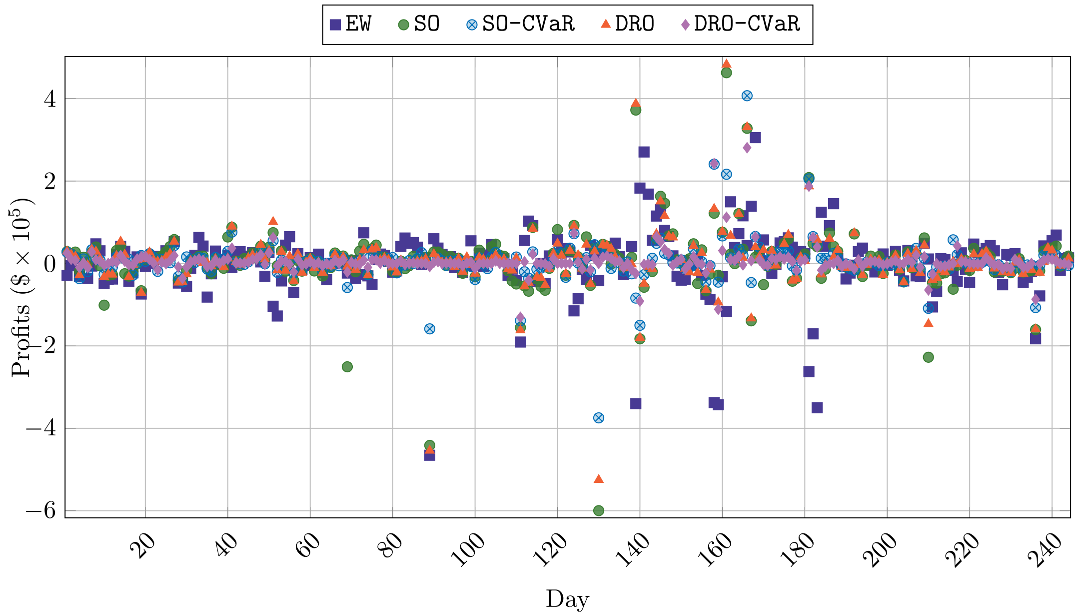
5 Numerical example

This section presents and analyzes the results of our strategy based on `DRO-CVaR`. The strategy is tested on a 8-month period with hyperparameters obtained from the tuning process introduced in Section 4.3. We remark that the 8 months of data utilized for testing was not used for hyperparameter tuning and, as such, is considered as out-of-sample.

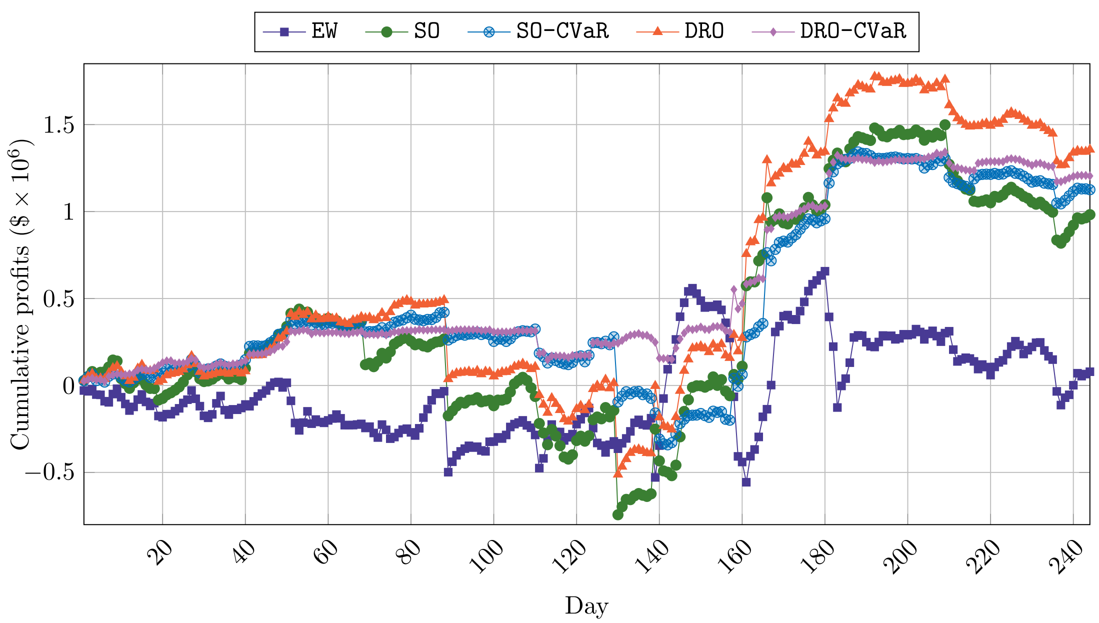
The case study is based on data from NYISO, which operates an electricity market with 11 geographic zones. These zones are aggregated from a larger number of buses, each representing a point of generation or consumption. Virtual bids are submitted for each hour before the virtual bidding market closing time (5 AM), and the system clears the DAM a day prior to the RTM. The case study considers placing bids for every hour of every day over an 8-month testing period, from February 1$^{\text{st}}$, 2024, to October 1$^{\text{st}}$, 2024.

We benchmark our approach against similar online virtual bidding strategies that substitute the `DRO-CVaR` model for the `SO`, `SO-CVaR`, and `DRO` models. Section 4.3 tuning procedure is used independently on all resulting strategies. We also include the equally-weighted (`EW`) strategy, which sells the same energy quantity across all buses i and hours of the day, subject to the same quantity constraints as the other models. An hourly quantity limit $L = 400$ MWh is used throughout this section. The intuition behind the strategy relies on the conservative approach typically adopted by ISOs. ISOs often purchase more electricity than required in the DAM to avoid any shortfalls in the RTM. The `EW` strategy takes advantage of this by assuming that the DAM price will generally be higher than the RTM price without focusing on specific buses. This commonly used benchmark [12] serves as a baseline and evaluates the performance of our strategy against a static one. Figure 3a presents the daily profits of every strategy on the 8-month testing period. The cumulative profits then are illustrated in Figure 3b.

Analyzing Figures 3a and 3b, it can be observed that combining distributionally robust optimization with CVaR tends to achieve both profitability and risk mitigation. Strategies employing the CVaR, namely `SO-CVaR` and `DRO-CVaR`, show reduced downside risk, although only `DRO-CVaR` consistently maintains positive profits. The `DRO`-based strategy (without CVaR) achieves the highest cumulative profit. We observe that `DRO-CVaR` never suffers the largest daily loss and is regularly amongst the most important earners. This indicates that the strategy is more robust and reliable compared to the benchmarks.The `DRO-CVaR`-based strategy has the most consistent profits, as losses are limited, leading to large cumulative profits. It is the only strategy that never has its cumulative profits dip below zero, which demonstrates better robustness than the other strategies. This comparison also illustrates that all data-driven strategies outperform the static benchmark by a significant margin.



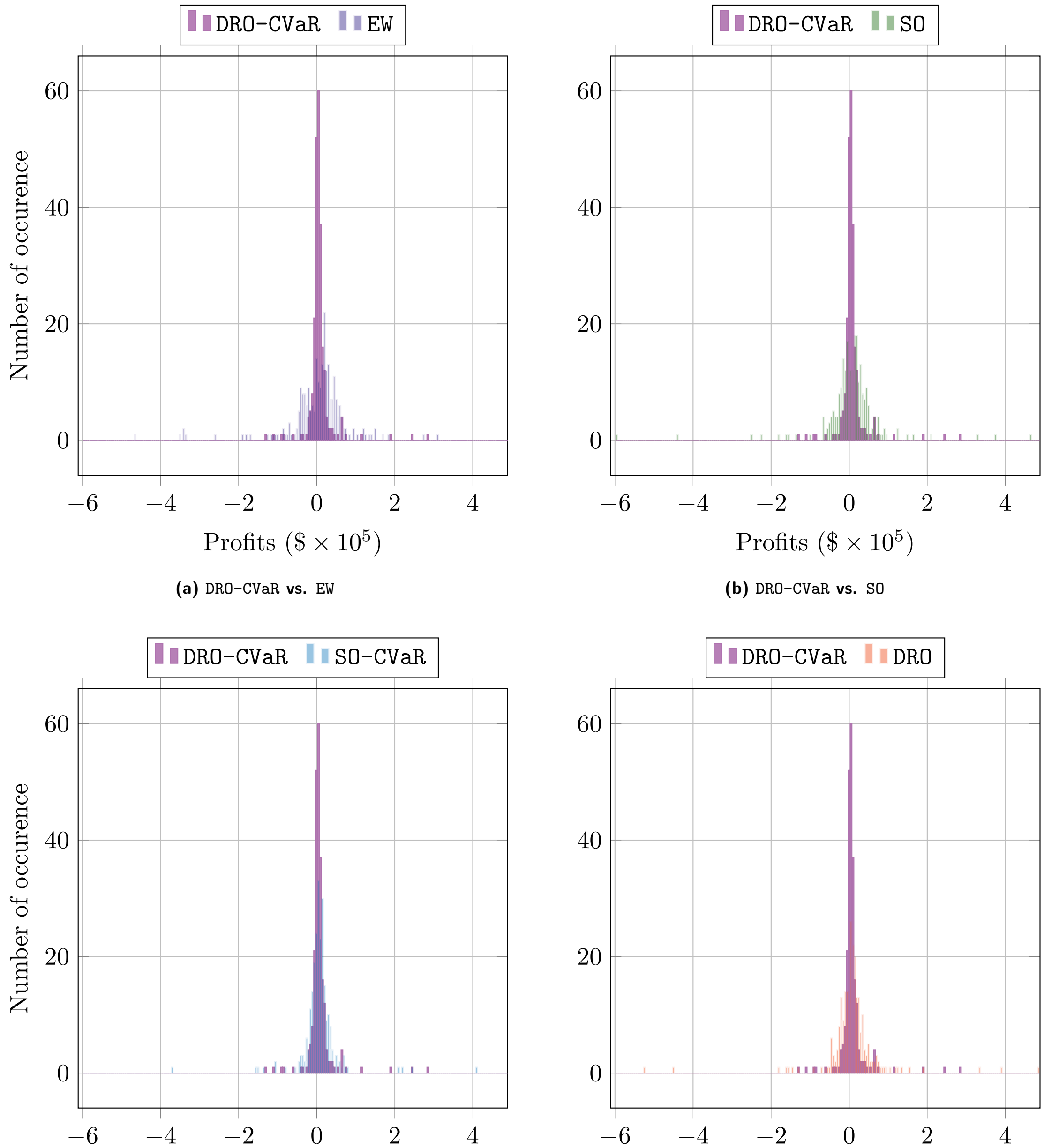
(a) Daily profits over the 8-month testing set



(b) Cumulative profits over the 8-month testing set

Figure 3: Profits over the 8-month testing set

Figure 4 illustrates the distribution of the daily profits of our strategy compared to all the benchmarks. We note that our strategy is less volatile than the other observed strategies. They are all centred slightly right of zero, which indicates profitability, but the main difference is that the number of negative profit occurrences is diminished, although at the cost of missing out higher positive profit opportunities.



(a) DRO-CVaR **vs.** EW

(b) DRO-CVaR **vs.** SO

(c) DRO-CVaR **vs.** SO-CVaR

(d) DRO-CVaR **vs.** DRO

Figure 4: Comparison of daily profits over the 8-month testing set

Figure 5 illustrates the distribution of profits across the different strategies, presenting the interquartile ranges, medians, whiskers, and outliers. It can be observed that the CVaR-based strategies exhibit reduced profit volatility, as evidenced by the narrower interquartile ranges and shorter whiskers. Among them, `DRO-CVaR` yields the lowest volatility. In evaluating the performance of these strategies, it is important to consider not just their total profits but also the profits relative to the quantity of bids placed. Each strategy operates under a consistent set of bidding constraints; however, some pursue a more aggressive bidding approach, bidding the maximum quantity at every timestep, while others adopt a more conservative strategy, opting not to bid to the limit. Figure 6 presents the scaled profits for each strategy, which are calculated as the cumulative profit at the end of the 8-month period divided by the total quantity of virtual bids. This measure provides insight into the profitability per MWh, illustrating how effectively each strategy translates bid quantities into realized gain.

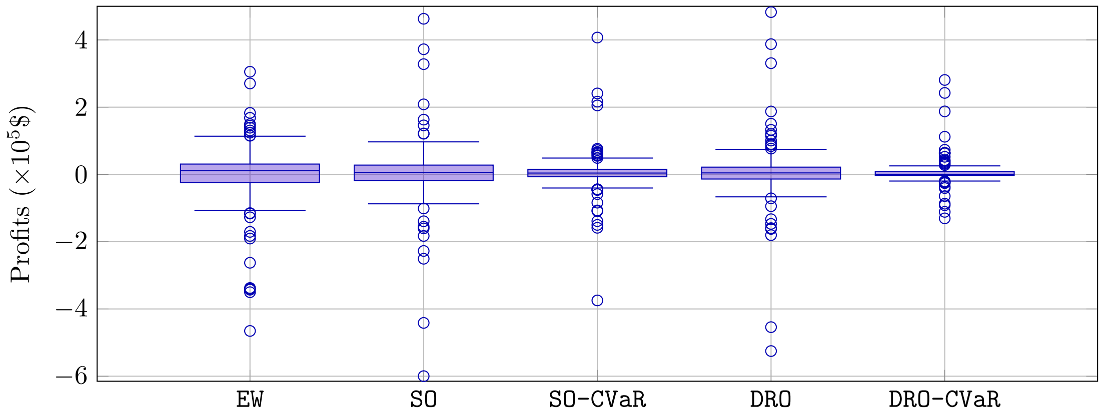


Figure 5: Boxplot of daily profits over the 8-month testing period

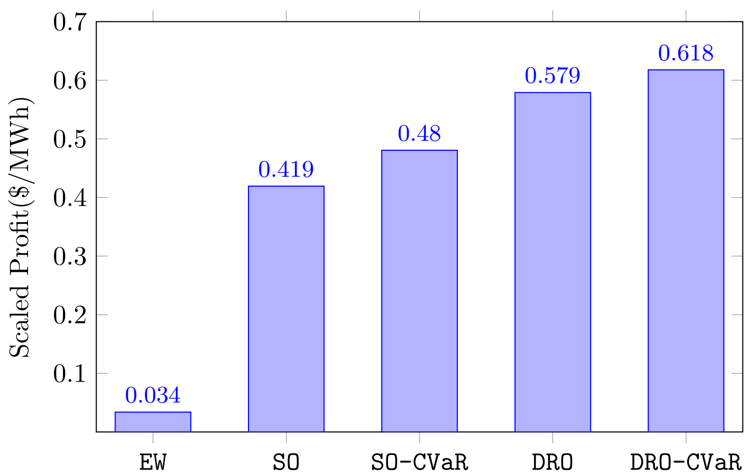


Figure 6: Scaled profit over the 8-month testing period

Figure 6 shows that the best performing strategy in terms of scaled profit is the one based on `DRO-CVaR`. The `DRO`-based strategy also demonstrates strong scaled profit, highlighting the importance of uncertainty considerations in increasing profits per MWh. The moderate scaled profits of `SO-CVaR` suggest it could be improved with additional robustness features while the very low cumulative gain-per-MWh of `EW` reflects a poor performance of the strategy.

To evaluate performance in a risk-adjusted context, two other metrics are used: the Sharpe and Calmar ratios. The Sharpe ratio measures the return earned per unit of risk, allowing us to assess how each strategy balances reward against volatility.

To calculate the Sharpe ratio S, we start with an initial portfolio value of $v_0 = 10^6$. The Sharpe ratio is then calculated as the mean scaled returns $\bar{\eta}$ over the total number of trading days $J = 244$ days (8 months) and the standard deviation σ_η of the scaled daily returns η_j as defined in (13) :

$$S = \frac{\bar{\eta}}{\sigma_\eta}\sqrt{J}.$$

The Sharpe ratio allows us to compare the reward versus risk consistently across strategies and periods [2]. A higher Sharpe ratio indicates a more efficient strategy as it achieves greater returns per unit of risk. In Figure 7, we illustrate the Sharpe ratios for each strategy.

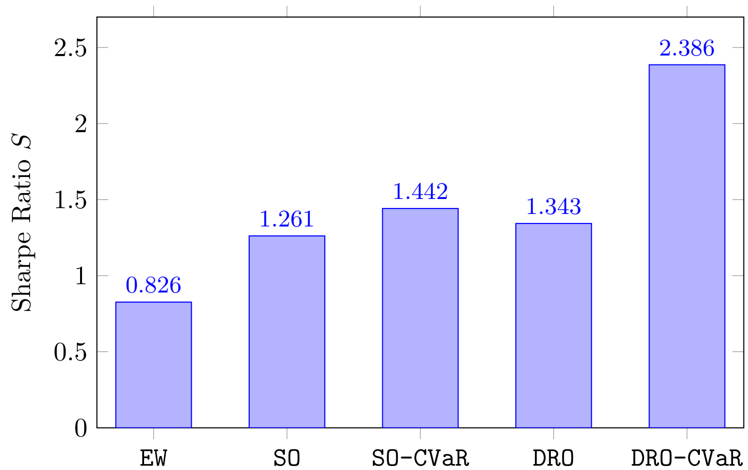


Figure 7: Sharpe ratios over the 8-month testing period

Figure 7 shows that the `DRO-CVaR` strategy significantly outperforms the other benchmarks, achieving a Sharpe ratio of 2.386. This indicates a highly efficient strategy with greater returns per unit of risk. Figure 7 validates that strategies incorporating CVaR achieve higher risk-return outcomes with `SO-CVaR` having the second best Sharpe ratio with 1.442. This analysis underscores the effectiveness of `DRO-CVaR` in balancing risk and profits.

Figure 8 provides Calmar ratios defined in (16), which assess the performance relative to the MDD for all strategies. A higher Calmar ratio indicates a strategy's ability to deliver returns that sufficiently compensate for its worst historical losses.

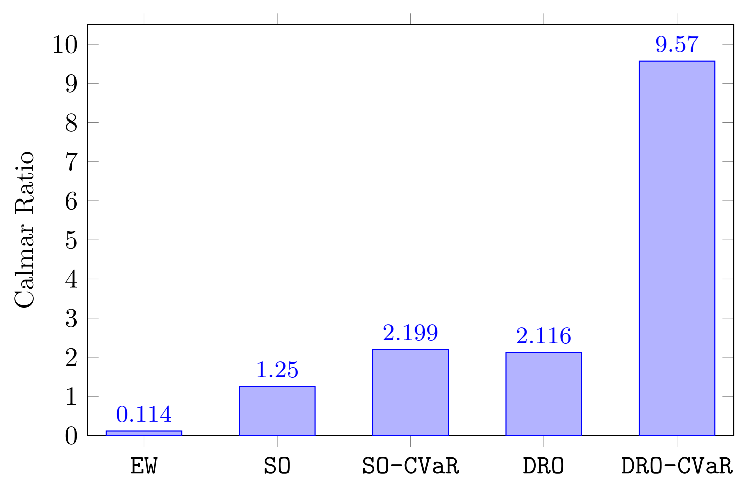


Figure 8: Calmar ratios over the 8-month testing period

The Calmar ratio analysis highlights the superior performance of the `DRO- CVaR` strategy, which achieves a remarkably high Calmar ratio of 9.570 as shown in Figure 8. In comparison, the `SO-CVaR` and `DRO`-based strategies have Calmar ratios of 2.199 and 2.116, respectively, indicating good performance though limited compared to `DRO-CVaR`. This reaffirms that our strategy is the top-performing one in terms of risk-adjusted returns.

The results of Figures 3b, 6, 7, and 8 are summarized in Table 2. The best values are displayed in bold characters.

Table 2: Performance metrics per strategies on the 8-month testing period

Metric	EW	SO	SO-CVaR	DRO	DRO-CVaR
Cumulative Profit($)	79,167	982,299	1,125,487	**1,356,383**	1,204,584
Scaled Profit($/MWh)	0.034	0.419	0.48	0.579	**0.618**
Sharpe Ratio	0.826	1.261	1.442	1.343	**2.386**
Calmar Ratio	0.114	1.250	2.199	2.116	**9.570**

6 Conclusion

In this work, we introduce a DRO approach to virtual bidding in two-settlement electricity markets. We begin by constructing a virtual bidding strategy using CVaR to mitigate risk. We then use the DRO framework to handle price uncertainty. The resulting `DRO-CVaR` model is formulated as a convex program and can be efficiently solved to optimality. The model is then integrated into an online strategy to compute daily bid quantities based on updated datasets of similar days, which are themselves selected from the last two years of data preceding each bidding day.

We propose a tuning approach to optimize hyperparameters, such as the dataset size and the Wasserstein ball radius, using historical data. Our strategy is backtested by tuning hyperparameters on 12 months of NYISO data and tested on an 8-month period. We benchmark our approach against four other strategies (`SO`, `SO-CVaR`, `DRO`-based and `EW`), achieving second place in cumulative profits and first in scaled cumulative profits, Sharpe ratio, and Calmar ratio.

For future work, we aim to build a linearized optimal power flow (DC-OPF) model to estimate the cleared price in the day-ahead market. By incorporating this model into our existing framework, we hope to make more accurate bids. Finally, we intend to extend our backtesting to include more diverse datasets, such as other markets, to validate the generalizability, and the effectiveness of our approach.

Appendix

We show that the box uncertainty set used to obtain (12) is a special case of the more general cone uncertainty set used in [24, (27)]. For the sequel, we only consider the variables and the parameters involved in the constraints that differ between (12) and [24, (27)]. The cone uncertainty set presented in [24, (27)] is defined by:

$$\mathcal{U} = \{\mathbf{s}_t \in \mathbb{R}^{|\mathcal{N}|} : \mathbf{C}^\top \mathbf{s}_t \leq \mathbf{d}, \quad \forall i \in \mathcal{N}, t \in \mathcal{T}\}. \tag{17}$$

The box uncertainty set of (12) is defined by:

$$\begin{aligned}\mathcal{S} &= \{\mathbf{s}_t \in \mathbb{R}^{|\mathcal{N}|} : |s_{i,t}| \leq \Lambda, \quad \forall i \in \mathcal{N}, t \in \mathcal{T}\} \\ &= \{\mathbf{s}_t \in \mathbb{R}^{|\mathcal{N}|} : s_{i,t} \leq \Lambda \text{ and } -s_{i,t} \leq \Lambda \quad \forall i \in \mathcal{N}, t \in \mathcal{T}\},\end{aligned}$$

where each component of $\mathbf{s}_t$, $s_{i,t}$, is bounded by a maximum value $\Lambda > 0$, ensuring that all elements of $\mathbf{s}_t$ remain within this fixed range. This is a special case of (17) where:

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} \\ -\mathbf{I} \end{pmatrix}, \qquad \mathbf{d} = \Lambda \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix}.$$

Let $\boldsymbol{\gamma} \in \mathbb{R}^{2|\mathcal{N}|}$. From [24, (27c)], we have :

$$\boldsymbol{\gamma}^\top (\mathbf{d} - \mathbf{C}\mathbf{s}_t) \geq 0$$

$$\Leftrightarrow -\boldsymbol{\gamma}^\top \begin{pmatrix} \mathbf{I} \\ -\mathbf{I} \end{pmatrix} \mathbf{s}_t + \boldsymbol{\gamma}^\top \Lambda \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix} \geq 0.$$

This allows us to re-express the last two terms of (12b) in terms of:

$$\boldsymbol{\nu} = -\begin{pmatrix} \mathbf{I} \\ -\mathbf{I} \end{pmatrix}^\top \boldsymbol{\gamma} = \begin{pmatrix} -\mathbf{I} & \mathbf{I} \end{pmatrix} \boldsymbol{\gamma}, \quad \mu = \boldsymbol{\gamma}^\top \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix}.$$

Let

$$\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix},$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{N}|}$ and $\boldsymbol{\beta} \in \mathbb{R}^{|\mathcal{N}|}$. We can rewrite $\boldsymbol{\nu}$ and μ: as$\boldsymbol{\nu} = -\boldsymbol{\alpha} + \boldsymbol{\beta}$ and $\mu = \mathbf{1}^\top(\boldsymbol{\alpha} + \boldsymbol{\beta})$. Recalling (12d) and substituting the above definitions, we obtain:

$$\begin{aligned} \mu &\geq \|\boldsymbol{\nu}\|_1 \\ \Leftrightarrow \mathbf{1}^\top(\boldsymbol{\alpha} + \boldsymbol{\beta}) &\geq |-\boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1| + |-\boldsymbol{\alpha}_2 + \boldsymbol{\beta}_2| + \ldots + |-\boldsymbol{\alpha}_n + \boldsymbol{\beta}_n| \\ \Leftrightarrow \sum_{i \in \mathcal{N}} (\boldsymbol{\alpha}_i + \boldsymbol{\beta}_i) &\geq \sum_{i \in \mathcal{N}} |\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i|. \end{aligned}$$

This is always true if $\boldsymbol{\gamma} \geq 0$, which coincides with with [24, (27e)]. Finally, we note that (12d):

$$\begin{aligned} \|\boldsymbol{\nu}\|_2 &\leq 0 \\ \Leftrightarrow \|-\boldsymbol{\alpha} + \boldsymbol{\beta}\|_2 &\leq 0 \\ \Leftrightarrow \|\mathbf{C}^\top \boldsymbol{\gamma}\|_2 &\leq 0, \end{aligned}$$

is equivalent to [24, (27d)]. Using the box uncertainty set in (12) is therefore a special case of the more general cone uncertainty set used in [24, (27)].

References

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.

[2] Sevi Baltaoglu, Lang Tong, and Qing Zhao. Algorithmic bidding for virtual trading in electricity markets. IEEE Transactions on Power Systems, 34(1):535–543, 2018.

[3] California Independent System Operator. California ISO, 2024. Accessed: 2024-10-14.

[4] Metin Celebi, Attila Hajos, and Philip Q Hanser. Virtual bidding: the good, the bad and the ugly. The Electricity Journal, 23(5):16–25, 2010.

[5] Ruidi Chen and Ioannis Ch Paschalidis. Distributionally robust learning. Foundations and Trends® in Optimization, 4(1-2):1–243, 2020.

[6] Jaehyung Choi, Hyangju Kim, and Young Shin Kim. Diversified reward-risk parity in portfolio construction. Studies in Nonlinear Dynamics & Econometrics, 2024.

[7] Ningning Du, Yankui Liu, and Ying Liu. A new data-driven distributionally robust portfolio optimization method based on Wasserstein ambiguity set. IEEE Access, 9:3174–3194, 2020.

[8] Electric Reliability Council of Texas. ERCOT, 2024. Accessed: 2024-10-14.

[9] Empyrical Library Documentation. Empyrical: Common financial risk and performance metrics, 2014. Accessed: 2024-09-12.

[10] National Grid ESO. National Grid ESO: Electricity Market Operations in the UK, 2025. Accessed: 2025-02-27.

[11] Rémi Galarneau-Vincent, Geneviève Gauthier, and Frédéric Godin. Foreseeing the worst: Forecasting electricity DART spikes. Energy Economics, 119:106521, 2023.

[12] Matteo Gelmini and Pierpaolo Uberti. The equally weighted portfolio still remains a challenging benchmark. International Economics, 179:100525, 2024.

[13] William W Hogan. Virtual bidding and electricity market design. The Electricity Journal, 29(5):33–47, 2016.

[14] ISO New England. ISO-NE, 2024. Accessed: 2024-10-14.

[15] Rajnish Kamat and Shmuel S Oren. Multi-settlement systems for electricity markets: Zonal aggregation under network uncertainty and market power. In Proceedings of the 35th Annual Hawaii International Conference on System Sciences, pages 739–748. IEEE, 2002.

[16] Jalal Kazempour and Benjamin F Hobbs. Value of flexible resources, virtual bidding, and self-scheduling in two-settlement electricity markets with wind generation—Part I: Principles and competitive model. IEEE Transactions on Power Systems, 33(1):749–759, 2017.

[17] Simge Küçükyavuz and Ruiwei Jiang. Chance-constrained optimization under limited distributional information: A review of reformulations based on sampling and distributional robustness. EURO Journal on Computational Optimization, 10:100030, 2022.

[18] Yinglun Li. Data-Driven Modeling and Algorithmic Trading in Electricity Market. PhD thesis, UC Riverside, 2024.

[19] Adriana A Londoño and Juan D Velásquez. Risk management in electricity markets: Dominant topics and research trends. Risks, 11(7):116, 2023.

[20] Malik Magdon-Ismail and Amir F Atiya. Maximum drawdown. Risk Magazine, 17(10):99–102, 2004.

[21] Hossein Mehdipourpicha, Siyuan Wang, and Rui Bo. Developing robust bidding strategy for virtual bidders in day-ahead electricity markets. IEEE Open Access Journal of Power and Energy, 8:329–340, 2021.

[22] Ali Mehdizadeh and Navid Taghizadegan. Robust optimisation approach for bidding strategy of renewable generation-based microgrid under demand side management. IET Renewable Power Generation, 11(11):1446–1455, 2017.

[23] Midcontinent Independent System Operator. MISO, 2024. Accessed: 2024-10-14.

[24] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming, 171(1):115–166, 2018.

[25] Antonio Naimoli and Giuseppe Storti. Forecasting volatility and tail risk in electricity markets. Journal of Risk and Financial Management, 14(7):294, 2021.

[26] New York Independent System Operator. NYISO, 2024. Accessed: 2024-10-14.

[27] OMI. Iberian Peninsula Electricity Market: OMI Overview, 2025. Accessed: 2025-02-27.

[28] Julien Pallage and Antoine Lesage-Landry. Wasserstein distributionally robust shallow convex neural networks. arXiv preprint arXiv:2407.16800, 2024.

[29] Pierre Pinson. Distributionally robust trading strategies for renewable energy producers. IEEE Transactions on Energy Markets, Policy and Regulation, 1(1):37–47, 2023.

[30] PJM Interconnection LLC. PJM, 2024. Accessed: 2024-10-14.

[31] R. Tyrell Rockafellar and Stanislav Uryasev. Optimization of conditional Value-at-Risk. Journal of Risk, 2:21–42, 2000.

[32] Ehsan Samani, Mahdi Kohansal, and Hamed Mohsenian-Rad. A data-driven convergence bidding strategy based on reverse engineering of market participants' performance: A case of California ISO. IEEE Transactions on Power Systems, 37(3):2122–2136, 2021.

[33] Southwest Power Pool. SPP, 2024. Accessed: 2024-10-14.

[34] Dirk Tasche. Expected shortfall and beyond. Journal of Banking & Finance, 26(7):1519–1533, 2002.