Online Appendix

The appendix is organized as follows. Appendix A details the calibration methods for all parameters using data in OU-DRO. Appendix B discusses OU distributions defined on the whole space $\mathbb{R}^d$ as the generalization of OU distributions defined on $\mathcal{D}_0$. Appendix C provides the details of POU distributions and POU-DRO discussed in Section 7. Appendix D contains the proofs of all our results.

Appendix A: Statistical Calibration of Constraints. To illustrate how to calibrate the constraints of formulation (13) with data, we consider the following example used in Section 8.2:

$$\begin{aligned}
\max\ & P((X,Y)\in S)\\
\text{subject to } & l_{\bar{F}} \le \bar{F}(x_0,y_0) \le u_{\bar{F}}\\
& f_X(x_0) \le u_X\\
& f_Y(y_0) \le u_Y\\
& a_{X,i} \le P(x_0 \le X \le x_i, Y \ge y_0 | X \ge x_0, Y \ge y_0) \le b_{X,i}, i=1,\dots,n_X,\\
& a_{Y,j} \le P(X \ge x_0, y_0 \le Y \le y_j | X \ge x_0, Y \ge y_0) \le b_{Y,j}, j=1,\dots,n_Y,\\
& f(x',y') \ge f(x,y) \text{ if } x_0 \le x' \le x \text{ and } y_0 \le y' \le y.
\end{aligned} \tag{40}$$

Here we do not include the lower bound density constraints since we see in Section 5.2 that they are redundant. Suppose we have m samples $(X_i,Y_i), i=1,\dots,m$ and the confidence level is set to be α. Suppose also that x_0 and y_0 are set as some coordinate-wise top percentiles of the data (e.g., top 10 percentiles) in order to locate the tail region. Then the constants in problem (40) can be calibrated

as

$$
\begin{aligned}
&l_{\bar{F}} = \hat{\bar{F}}(x_0,y_0) - \frac{\Phi^{-1}(1-\alpha/14)}{\sqrt{m}}\hat{\sigma}_{\bar{F}(x_0,y_0)}, \quad u_{\bar{F}} = \hat{\bar{F}}(x_0,y_0) + \frac{\Phi^{-1}(1-\alpha/14)}{\sqrt{m}}\hat{\sigma}_{\bar{F}(x_0,y_0)},\\
&u_X = \hat{f}_{1-\alpha/7,X|Y\geq y_0}(x_0)\left(\hat{P}(Y\geq y_0) + \frac{\Phi^{-1}(1-\alpha/7)}{\sqrt{m}}\hat{\sigma}_{P(Y\geq y_0)}\right),\\
&u_Y = \hat{f}_{1-\alpha/7,Y|X\geq x_0}(y_0)\left(\hat{P}(X\geq x_0) + \frac{\Phi^{-1}(1-\alpha/7)}{\sqrt{m}}\hat{\sigma}_{P(X\geq x_0)}\right),\\
&a_{X,i} = \hat{F}_X(x_i) - \frac{q}{\sqrt{m_1}}, \quad b_{X,i} = \hat{F}_X(x_i-) + \frac{q}{\sqrt{m_1}},\\
&a_{Y,j} = \hat{F}_Y(y_j) - \frac{q}{\sqrt{m_1}}, \quad b_{Y,j} = \hat{F}_Y(y_j-) + \frac{q}{\sqrt{m_1}}.
\end{aligned}
$$

We will explain each notation and provide justifications. First, we have

$$\hat{\bar{F}}(x_0,y_0) = \frac{1}{m}\sum_{i=1}^{m} I(X_i\geq x_0, Y_i\geq y_0),$$

$$\hat{\sigma}_{\bar{F}(x_0,y_0)} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(I(X_i\geq x_0,Y_i\geq y_0)-\hat{\bar{F}}(x_0,y_0))^2}$$

as the point estimate of $\bar{F}(x_0,y_0)$ given by the sample mean of $I(X_i\geq x_0,Y_i\geq y_0)$'s, and the sample standard deviation of $I(X_i\geq x_0,Y_i\geq y_0)$'s respectively. We also denote Φ as the cumulative distribution function of the standard normal distribution. Thus, the constants $l_{\bar{F}}$ and $u_{\bar{F}}$ are set as the asymptotically exact lower and upper confidence bounds of $\bar{F}(x_0,y_0)$ induced by the standard central limit theorem (CLT) at the confidence level $1-\alpha/7$. Here, the confidence level $1-\alpha/7$ is due to a Bonferroni correction which we will explain later.

Next, the constant u_X is calibrated as the $(1-\alpha/7)$-level upper confidence bound on $f_X(x_0)$ by the bootstrap as follows. Note that the truncated marginal density $f_X(x_0)$ can be written as $f_X(x_0) = P(Y\geq y_0)f_{X|Y\geq y_0}(x_0)$, where $f_{X|Y\geq y_0}$ is the density of X conditional on $\{Y\geq y_0\}$. We need to find the asymptotically exact $(1-\alpha/7)$-level upper bound for both factors. For $f_{X|Y\geq y_0}(x_0)$, we select the data (X_i,Y_i) with $Y_i\geq y_0$. Suppose the number of such data is m_0. We resample the x-coordinates of the selected data with replacement to the size m_0, and use the resample to construct a kernel density estimate for X at x_0. Repeat the procedure many times and the $(1-\alpha/7)$-th quantile $\hat{f}_{1-\alpha/7,X|Y\geq y_0}(x_0)$ of these estimates is an asymptotically exact $(1-\alpha/7)$-level upper bound

of $f_{X|Y\geq y_0}(x_0)$ [23]. On the other hand, an asymptotically exact $(1-\alpha/7)$-level upper confidence bound of $P(Y \geq y_0)$ is obtained by the CLT like for $\bar{F}(x_0, y_0)$ discussed earlier, where

$$\hat{P}(Y \geq y_0) = \frac{1}{m}\sum_{i=1}^{m} I(Y_i \geq y_0), \quad \hat{\sigma}_{P(Y\geq y_0)} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(I(Y_i \geq y_0) - \hat{P}(Y \geq y_0))^2}.$$

The calibration of u_Y is justified similarly.

Furthermore, to calibrate the constants $a_{X,i}, b_{X,i}$, we use the Kolmogorov–Smirnov statistic. We collect all data such that $X \geq x_0$ and $Y \geq y_0$. Suppose these data have size m_1. The Kolmogorov–Smirnov test gives us

$$\sqrt{m_1}\sup_{x\geq x_0}|F_X(x) - \hat{F}_X(x)| \leq q \tag{41}$$

an asymptotically exact $(1-\alpha/7)$-level confidence region on $F_X(x)$, where $F_X(x) = P(X \leq x | X \geq x_0, Y \geq y_0)$ and $\hat{F}_X$ is the empirical distribution of X restricted to $X \geq x_0, Y \geq y_0$, and q is the $(1-\alpha/7)$-th quantile of $\sup_{t\in[0,1]}|BB(t)|$ where BB is the standard Brownian bridge. By considering the left and right limits at $x = x_i$ in (41), we can rewrite (41) to obtain the lower and upper bounds of $F_X(x_i)$ as $a_{x,i}$ and $b_{X,i}$ defined earlier, where $\hat{F}_X(x_i-)$ is the left limit of $\hat{F}_X$ at x_i. The constants $a_{Y,j}, b_{Y,j}$ can be calibrated similarly.

Finally, the Bonferroni correction is applied since there are in total 7 sets of parameters that we need to calibrate, and we want to control the familywise Type I error to be at most α, thus giving each individual confidence level discussed above to be $1-\alpha/7$. If the constraints are calibrated via the methods above and the true density is OU about (x_0, y_0) on $[x_0, \infty) \times [y_0, \infty)$, then the constraints in (40) will hold simultaneously with confidence level $1-\alpha$ asymptotically. By Corollary 1, the optimal value will then be an asymptotically valid upper confidence bound on the true probability with level $1-\alpha$.

Appendix B: OU Distribution on $\mathbb{R}^d$. In this section, we consider a generalization of the OU distribution defined in Definition 6. We will enlarge the support from $\mathcal{D}_0$ to $\mathbb{R}^d$ and keep the monotonicity property for each variable. Like the OU distribution on $\mathcal{D}_0$, we will provide two definitions for the OU distribution on $\mathbb{R}^d$. One is defined in terms of the density and the other is defined via mixture representation.

Definition 7 (Orthounimodal density on $\mathbb{R}^d$). A probability distribution on $\mathbb{R}^d$ with density f (with respect to the Lebesgue measure) is OU about mode $x_0 = (x_{10}, \ldots, x_{d0})$ if for each $i = 1, \ldots, d$ and any fixed $x_j \in \mathbb{R}, j \neq i$, the function $x_i \mapsto f(x)$ is non-decreasing on $(-\infty, x_{i0}]$ and non-increasing on $[x_{i0}, \infty)$. □

DEFINITION 8 (MIXTURE REPRESENTATION OF OU DISTRIBUTION ON $\mathbb{R}^d$). We have:

- A set $K \subset \mathbb{R}^d$ is said to be orthounimodal about x_0 if for every $x \in K$, the closed rectangle with edges parallel to the axes and with opposite vertices x_0 and x is in K, i.e., $(\eta_1 x_1 + (1-\eta_1)x_{10}, \ldots, \eta_d x_d + (1-\eta_d)x_{d0}) \in K$ for any $\eta_i \in [0,1], i = 1, \ldots, d$.
- A distribution on $\mathbb{R}^d$ is called OU about x_0 if it belongs to the closed convex hull of the set of all uniform distributions on subsets of $\mathbb{R}^d$ which are OU about x_0. □

From Definition 7 and Definition 8, we can see OU distribution has a linear transformation property: if X is OU about x_0, then $X + x_0'$ is OU about $x_0 + x_0'$. Besides, analogies of Lemma 2, Theorem 1 and Lemma 3 hold for OU distributions on $\mathbb{R}^d$.

LEMMA 9. *Suppose that $K \subset \mathbb{R}^d$ is an OU set about x_0. Then K is Lebesgue measurable. Besides, $\bar{K}$ and $\bar{K^\circ}$ (closure of K°) are also OU sets about x_0 satisfying $\lambda(K^\circ) = \lambda(K) = \lambda(\bar{K})$.*

The following theorem justifies the equivalence of Definition 8 and Definition 7 and establishes the Choquet representation for OU distributions on $\mathbb{R}^d$ in the presence of density. To simplify the notations, we assume without loss of generality that the mode is the origin. If X is OU about x_0, we can see $X - x_0$ is OU about the origin and Theorem 3 applies. Recall that W_K denotes the uniform distribution on K.

THEOREM 3. *Suppose a distribution P on $\mathbb{R}^d$ is absolutely continuous with respect to Lebesgue measure. Then P is OU about the origin if and only if there is a density $f(x)$ of P such that for every $s > 0$, the set*

$$C_s = \{x \in \mathcal{D}_0 : f(x) \geq s\}$$

is OU about the origin, or equivalently, if and only if f is an OU density about the origin on $\mathbb{R}^d$. Besides, let $O_1, \ldots, O_{2^d}$ be the 2^d (closed) orthants of $\mathbb{R}^d$. Then P has the following Choquet representation:

$$P(B) = \sum_{1 \leq i \leq 2^d, P(O_i) > 0} P(O_i) \int_0^\infty W_{\bar{C}_s^i}(B) g_i(s) ds \tag{42}$$

for any Lebesgue measurable set B, where $C_s^i = \{x \in O_i : f_i(x) \geq s\}$, $f_i(x) = f(x) I(x \in O_i)/P(O_i)$ is the density of the conditional distribution $P(\cdot|O_i)$, $\bar{C}_s^i$ is the closure of C_s^i and $g_i(s) = \lambda(\bar{C}_s^i)$ is a probability density on $(0, \infty)$.

The following Lemma tells us about the extreme point in the class of OU distributions about the origin on $\mathbb{R}^d$.

LEMMA 10. *If $K \subset O_i$ is an OU set about the origin with $\lambda(K) > 0$ for some orthant O_i, then W_K is an extreme point in the class of OU distributions about the origin on $\mathbb{R}^d$.*

The condition $K \subset O_i$ is essential for Lemma 10. Consider an example in the bivariate case. Define $K = \{(x, y) : -1 \le x \le 1, -1 \le y \le 1\}$, $K_1 = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1\}$, $K_2 = \{(x, y) : -1 \le x \le 0, 0 \le y \le 1\}$, $K_3 = \{(x, y) : -1 \le x \le 0, -1 \le y \le 0\}$ and $K_4 = \{(x, y) : 0 \le x \le 1, -1 \le y \le 0\}$. We can see all the sets are OU about the origin. However, W_K is not an extreme point in the class of OU distributions about the origin on $\mathbb{R}^2$ since it can be written as

$$W_K = \frac{1}{4}W_{K_1} + \frac{1}{4}W_{K_2} + \frac{1}{4}W_{K_3} + \frac{1}{4}W_{K_4}.$$

This explains why the Choquet representation in Theorem 3 is represented as (42) instead of

$$P(B) = \int_0^\infty W_{\tilde{C}_s}(B)g(s)ds \tag{43}$$

although (43) still holds in the setting of Theorem 3.

Appendix C: Theory of Partial Orthounimodality and Partially Orthounimodal Distributionally Robust Optimization. In this section, we provide the details on the theory of POU distributions and the POU-DRO problem discussed in Section 7.

C.1. POU Distribution. In this section, we introduce POU and establish its Choquet representation by following the same flow as in the OU case. As we have discussed, POU models a density that is non-increasing in only some of the dimensions. For simplicity, we suppose the density is non-increasing in only the first d' components. In other words, only the first d' components of the random vector are in their tails. For other cases, we can simply permute the components to transform them into this case. Let $\mathcal{D}_0 = \{x \ge x_0\}$ where $x_0 = (x_{10}, \dots, x_{d0}) \in \mathbb{R}^{d'} \times [-\infty, \infty)^{d-d'}$ for some $d' < d$. When $x_{j0} = -\infty$, $x_j \ge x_{j0}$ is interpreted as $x_j \in \mathbb{R}$. Similar to OU distributions, POU distributions can be defined in terms of monotonicity of the probability density or mixture representation.

DEFINITION 9 (d'-PARTIALLY ORTHOUNIMODAL DENSITY). A probability distribution on $\mathcal{D}_0$ with density f (with respect to the Lebesgue measure) is d'-partially orthounimodal (d'-POU) about mode x_0 if f is non-increasing with respect to $x_1, \dots, x_{d'}$ on $\mathcal{D}_0$ for any fixed $(x_{d'+1}, \dots, x_d) \ge (x_{d'+1,0}, \dots, x_{d0})$. □

DEFINITION 10 (MIXTURE REPRESENTATION OF d'-POU DISTRIBUTION). We have:

- A measurable set $K \subset \mathcal{D}_0$ is said to be d'-POU about x_0 if for every $x \in K$, we have $x' \in K$ if $x_i \geq x_i' \geq x_{i0}$ for $i = 1, \ldots, d'$ and $x_i = x_i' \geq x_{i0}$ for $i = d'+1, \ldots, d$.
- A distribution on $\mathcal{D}_0$ is called d'-POU about x_0 if it belongs to the closed convex hull of the set of all uniform distributions on subsets of $\mathcal{D}_0$ which are d'-POU about x_0. □

The following theorem justifies the equivalence of Definitions 9 and 10 in the presence of a density.

THEOREM 4. *Suppose a distribution P on $\mathcal{D}_0$ is absolutely continuous with respect to Lebesgue measure. Then P is d'-POU about x_0 if and only if there is a density $f(x)$ of P such that for every $s > 0$, the set*

$$C_s = \{x \in \mathcal{D}_0 : f(x) \geq s\}$$

is d'-POU about x_0, or equivalently, if and only if f is a d'-POU density about x_0 on $\mathcal{D}_0$.

The Choquet representation of d'-POU distributions is slightly more complicated than that of the OU distributions. We first introduce the notations. For a d'-POU set K about x_0 on $\mathcal{D}_0$, we define its slice with respect to the subspace $\{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1}, \ldots, y_d = x_d\}$ as $K_{x_{d'+1},\ldots,x_d} = \{(y_1, \ldots, y_{d'}) : (y_1, \ldots, y_{d'}, x_{d'+1}, \ldots, x_d) \in K\} \subset \mathbb{R}^{d'}$. In other words, $K \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1}, \ldots, y_d = x_d\} = K_{x_{d'+1},\ldots,x_d} \times \{(x_{d'+1}, \ldots, x_d)\}$. Since K is d'-POU, $K_{x_{d'+1},\ldots,x_d}$ is either empty or a nonempty OU set about $(x_{10}, \ldots, x_{d'0})$ on $\{y \in \mathbb{R}^{d'} : y_i \geq x_{i0}, i = 1, \ldots, d'\}$. Recall that W_K denotes the uniform distribution on K and K could be in a lower-dimensional subspace of $\mathbb{R}^d$. In the presence of a density, we have the following Choquet representation theorem:

THEOREM 5. *Suppose a distribution P on $\mathcal{D}_0$ is absolutely continuous with respect to Lebesgue measure and P is d'-POU about x_0. Let f be a d'-POU density of P and $C_s = \{x \in \mathcal{D}_0 : f(x) \geq s\}$. Then the Choquet representation of P is given by:*

$$P(B) = \int_0^\infty \int_{x_{d'+1,0}}^\infty \cdots \int_{x_{d0}}^\infty W_{C_s \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1}, \ldots, y_d = x_d\}}(B) g(s, x_{d'+1}, \ldots, x_d) dx_{d'+1} \cdots dx_d ds$$

for any Lebesgue measurable set B, where $g(s, x_{d'+1}, \ldots, x_d) = \lambda_{d'}(C_{s,x_{d'+1},\ldots,x_d})$ is a probability density on $\{(s, x_{d'+1}, \ldots, x_d) : s > 0, x_i \geq x_{i0}, i = d'+1, \ldots, d\}$, $C_{s,x_{d'+1},\ldots,x_d}$ is the slice of C_s with respect to the subspace $\{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1}, \ldots, y_d = x_d\}$ and $\lambda_{d'}(\cdot)$ is the Lebesgue measure on $\mathbb{R}^{d'}$.

In the proof of Theorem 5, we need the following lemma to ensure the representation of P is indeed a Choquet representation.

LEMMA 11. *Suppose $K \subset \mathcal{D}_0$ is a d'-POU set about x_0. If K is fully contained in the subspace $\{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1}, \ldots, y_d = x_d\}$ for $x_i \geq x_{i0}, i = d'+1, \ldots, d$ with $\lambda_{d'}(K_{x_{d'+1},\ldots,x_d}) > 0$, then W_K is an extreme point in the class of d'-POU distributions about x_0 on $\mathcal{D}_0$.*

When we consider $d' = 1$, i.e., 1-POU distributions, we can see the slice $K_{x_2,\ldots,x_d}$ is either $[x_{10}, x_1)$ or $[x_{10}, x_1]$ for some $x_1 \geq x_{10}$. Such a simple form of $K_{x_2,\ldots,x_d}$ gives us a cleaner Choquet representation even without the presence of a density.

THEOREM 6. *Suppose a random vector $X = (X_1, \ldots, X_d)$ takes values on $\mathcal{D}_0$. Then the probability measure P of X is 1-POU about x_0 if and only if X is distributed as $(x_{10} + U(Z_1 - x_{10}), Z_2, \ldots, Z_d)$ where U is the uniform distribution on $(0,1)$ and independent of $(Z_1, \ldots, Z_d) \in \mathcal{D}_0$. Consequently, P is 1-POU about x_0 if and only if it has the following Choquet representation*

$$P = \int_{\mathcal{D}_0} W_{1\text{-}POU}(z) dQ(z), \tag{44}$$

where $W_{1\text{-}POU}(z)$ is the distribution of $(x_{10} + U(z_1 - x_{10}), z_2, \ldots, z_d)$, U is the uniform distribution on $(0,1)$ and Q is the distribution of $(Z_1, \ldots, Z_d)$ uniquely determined by P. Moreover, if $P(X_1 = x_{10}) = 0$, then X_1 is an absolutely continuous random variable with the probability density $E_Q[I(Z_1 \geq x)/(Z_1 - x_{10})], x \geq x_{10}$ which is continuous at $x = x_{10}$ in the sense of

$$\lim_{x \downarrow x_{10}} E_Q\left[\frac{I(Z_1 \geq x)}{Z_1 - x_{10}}\right] = E_Q\left[\frac{I(Z_1 \geq x_{10})}{Z_1 - x_{10}}\right].$$

C.2. POU-DRO Problem: Formulation and Reduction. In this section, we discuss the formulation of the d'-POU-DRO problem and the reduction of the 1-POU-DRO problem. With these, for the bivariate case, theories in this paper would cover all the extreme cases: if two dimensions are both in the tail, we can handle it by the OU-DRO problem; if only one dimension is in the tail, we can handle it by the 1-POU-DRO problem. Since the general d'-POU-DRO problem involves the OU-DRO problem (when $d' = d$) and the latter hasn't been solved completely for any dimension, we will leave the reduction of the general d'-POU-DRO problem as the future work.

Recall that for d'-POU, only the first d' dimensions are in the tail and the tail region is given by $\mathcal{D}_0 = \{x \geq x_0\}$ with $x_0 = (x_{10}, \ldots, x_{d0}) \in \mathbb{R}^{d'} \times [-\infty, \infty)^{d-d'}$. Imitating the OU-DRO problem, we formulate the d'-POU as follows:

$$\begin{aligned} \max \ & P((X_1, \ldots, X_d) \in S) \\ \text{subject to } & l_{\bar{F}} \leq \bar{F}(x_0) \leq u_{\bar{F}} \\ & f_{X_i}(x_{i0}) \leq u_{X_i}, i = 1, \ldots, d' \\ & a_i \bar{F}(x_0) \leq P(\underline{x}_{ji} \leq X_j \leq \bar{x}_{ji}, j = 1, \ldots, d) \leq b_i \bar{F}(x_0), i = 1, \ldots, n \\ & f(x') \geq f(x) \text{ if } x_i \geq x_i' \geq x_{i0}, i \leq d' \text{ and } x_i = x_i' \geq x_{i0}, i > d'. \end{aligned} \tag{45}$$

The interpretation and calibration of the constraints in (45) are the same as the OU-DRO problem (13). However, we do not restrict the truncated marginal densities for the last $d - d'$ components because these components are not in the tail in the motivating problems.

Now we focus on the reduction of the 1-POU-DRO problem. Here we make a small modification that we will use Definition 10 instead of Definition 9 as the 1-POU shape constraint. In other words, we will not assume the existence of the density of X. As we will see later in Theorem 7 and Corollary 4, this is because we may not be able to recover an absolutely continuous X for the 1-POU-DRO problem from the optimal solution to the reduced moment problem. As in the OU-DRO problem, we will replace $l_{\bar{F}} \leq \bar{F}(x_0) \leq u_{\bar{F}}$ by $\bar{F}(x_0) = c$ in the reduction procedure. For reference, the 1-POU-DRO problem is formulated as

$$\begin{aligned} \max \ & P((X_1, \ldots, X_d) \in S) \\ \text{subject to } & \bar{F}(x_0) = c \\ & f_{X_1}(x_{10}) \leq u_{X_1} \\ & a_i \bar{F}(x_0) \leq P(\underline{x}_{ji} \leq X_j \leq \bar{x}_{ji}, j = 1, \ldots, d) \leq b_i \bar{F}(x_0), i = 1, \ldots, n \\ & (X_1, \ldots, X_d) \text{ is 1-POU about } x_0 \text{ when restricted to } \mathcal{D}_0 \text{ and } P(X_1 = x_{10}) = 0. \end{aligned} \tag{46}$$

Here $f_{X_1}(x_1)$ denotes the marginal density of X_1 within $\mathcal{D}_0$, i.e., for any measurable $A \subset [x_{10}, \infty)$, $\int_A f_{X_1}(x_1) dx_1 = P(X_1 \in A, X_i \geq x_{i0}, i \geq 2)$. The existence of $f_{X_1}(x_1)$ is ensured by Theorem 6. We

require $f_{X_1}(x_1)$ to be continuous at $x_1 = x_{10}$ in order to avoid the arbitrariness of the density at this point. When X has a continuous density f satisfying Definition 9 on $\mathcal{D}_0$, f_{X_1} is just given by

$$f_{X_1}(x_1) = \int_{x_{20}}^{\infty} \cdots \int_{x_{d0}}^{\infty} f(x_1, \ldots, x_d) dx_2 \cdots dx_d$$

as in the formulation of OU-DRO problem. For the rare event set S, we assume it is in the form

$$S = \{x \in \mathcal{D}_0 : g_1(x_2, \ldots, x_d) \le x_1 \le g_2(x_2, \ldots, x_d)\} \tag{47}$$

where $g_i : [x_{20}, \infty) \times \cdots \times [x_{d0}, \infty) \mapsto [x_{10}, \infty]$ is a measurable function for $i = 1, 2$ and $g_1(x_2, \ldots, x_d) \le g_2(x_2, \ldots, x_d)$ for any $(x_2, \ldots, x_d)$. We can see the formulation (46) only depends on the behavior in $\mathcal{D}_0$ so we can restrict our attention to the truncated distribution on $\mathcal{D}_0$. The idea for solving (46) is to use the Choquet representation (44) to rewrite it as a moment problem with respect to the probability measure Q.

THEOREM 7. *The 1-POU-DRO problem (46) is equivalent to the following moment problem:*

$$\begin{aligned} &\max \ cE_Q\left[\frac{\min(g_2(Z_2, \ldots, Z_d), Z_1) - \min(g_1(Z_2, \ldots, Z_d), Z_1)}{Z_1 - x_{10}}\right] \\ &\textit{subject to } E_Q\left[\frac{1}{Z_1 - x_{10}}\right] \le \frac{u_{X_1}}{c} \\ &\qquad a_i \le E_Q\left[\frac{\min(Z_1, \bar{x}_{1i}) - \min(Z_1, \underline{x}_{1i})}{Z_1 - x_{10}} I(\underline{x}_{ji} \le Z_j \le \bar{x}_{ji}, j \ge 2)\right] \le b_i, i = 1, \ldots, n, \end{aligned} \tag{48}$$

where the decision variable Q is the probability distribution of $(Z_1, \ldots, Z_d) \in \mathcal{D}_0$. If $(Z_1^*, \ldots, Z_d^*) \sim Q^*$ is the optimal solution to the moment problem (48), then cP^* is the optimal solution to the 1-POU-DRO problem (46) where P^* is the distribution of $(x_{10} + U(Z_1^* - x_{10}), Z_2^*, \ldots, Z_d^*)$ with U being the uniform distribution on $(0, 1)$ and independent of $(Z_1^*, \ldots, Z_d^*)$.

According to Theorem 3.2 in [95], to solve the moment problem (48), it suffices to consider discrete probability measures with at most $n + 2$ points in the support. So (48) is equivalent to the

following non-linear optimization:

$$\begin{aligned}
\max\ & c\sum_{l=1}^{n+2} p_l \frac{\min(g_2(z_{l2},\ldots,z_{ld}),z_{l1}) - \min(g_1(z_{l2},\ldots,z_{ld}),z_{l1})}{z_{l1}-x_{10}} \\
\text{subject to}\ & \sum_{l=1}^{n+2} \frac{p_l}{z_{l1}-x_{10}} \le \frac{u_{X_1}}{c} \\
& a_i \le \sum_{l=1}^{n+2} p_l \frac{\min(z_{l1},\bar{x}_{1i}) - \min(z_{l1},\underline{x}_{1i})}{z_{l1}-x_{10}} I(\underline{x}_{ji} \le z_{lj} \le \bar{x}_{ji}, j \ge 2) \le b_i, i = 1,\ldots,n \qquad (49) \\
& \sum_{l=1}^{n+2} p_l = 1, p_l \ge 0, l = 1,\ldots,n+2 \\
& z_{li} \ge x_{i0},, l = 1,\ldots,n+2, i = 1,\ldots,d,
\end{aligned}$$

where p_l, z_{li} are the decision variables. The non-linear optimization (49) is capable to handle the 1-POU-DRO problem (46) when the constraint $\bar{F}(x_0) = c$ is replaced by $l_{\bar{F}} \le \bar{F}(x_0) \le u_{\bar{F}}$. In this case, we simply make c an additional decision variable varying in $[l_{\bar{F}}, u_{\bar{F}}]$. This is stated in the following corollary.

COROLLARY 4. *We have the following:*

(1) *The 1-POU-DRO problem (46), the moment problem (48) and the non-linear optimization problem (49) are equivalent.*

(2) *Consider the 1-POU-DRO problem (46) with* $\bar{F}(x_0) = c$ *replaced by* $l_{\bar{F}} \le \bar{F}(x_0) \le u_{\bar{F}}$. *Then this 1-POU-DRO problem is equivalent to the non-linear optimization problem (49) with an additional decision variable* c *and an additional constraint* $l_{\bar{F}} \le c \le u_{\bar{F}}$. *If* c^*, p_l^*, z_{li}^* *is the optimal solution to this non-linear optimization problem and we write* $(Z_1^*,\ldots,Z_d^*) \sim Q^*$ *as the discrete distribution on* $(z_{l1}^*,\ldots,z_{ld}^*), l = 1,\ldots,n+2$ *with corresponding probability* p_l^*, *then* c^*P^* *is the optimal solution to this 1-POU-DRO problem where* P^* *is the distribution of* $(x_{10} + U(Z_1^* - x_{10}), Z_2^*,\ldots,Z_d^*)$ *with* U *being the uniform distribution on* $(0,1)$ *and independent of* $(Z_1^*,\ldots,Z_d^*)$.

Corollary 4 reveals why we do not assume the existence of the density in the formulation of the 1-POU-DRO problem (46). Suppose p_l^*, z_{li}^* is the optimal solution to the non-linear optimization problem (49), i.e., the optimal solution to the moment problem (48) can be taken as the discrete distribution $(Z_1^*,\ldots,Z_d^*) \sim Q^*$ on $(z_{l1}^*,\ldots,z_{ld}^*), l = 1,\ldots,n+2$ with corresponding probability p_l^*. Then if we recover the optimal solution cP^* in the way described in Theorem 7, we will find cP^*

is not absolutely continuous on $\mathcal{D}_0$ as its support has Lebesgue measure 0. Due to this reason, we do not assume the existence of the density in (46).

We close this section by discussing a variant of formulation (46). As in the discussion of OU-DRO problem, we can replace the moment constraint in (46) by the unconditional version:

$$a_i \le P(\underline{x}_{ji} \le X_j \le \bar{x}_{ji}, j = 1, \dots, d) \le b_i. \tag{50}$$

In this case, we just need to replace a_i and b_i in the problem (48) or (49) by a_i/c or b_i/c and Theorem 7 and Corollary 4 still hold since (50) is equivalent to the conditional version constraint:

$$\frac{a_i}{c}\bar{F}(x_0) \le P(\underline{x}_{ji} \le X_j \le \bar{x}_{ji}, j = 1, \dots, d) \le \frac{b_i}{c}\bar{F}(x_0)$$

given $\bar{F}(x_0) = c$.

Appendix D: Proofs.

Proof of Lemma 1. The conclusion follows since

$$\left\{ \begin{array}{c} \text{geometric property of the distribution } F \text{ holds for } X \ge u, \\ \text{auxiliary constraints on } F \text{ holds} \end{array} \right\}$$

implies $\{Z^* \ge Z\}$. □

Proof of Proposition 1. We first prove the upper bound of the optimal value of (7). Consider any feasible density f. By changing into the polar coordinate, the objective function can be written as

$$P_f(0 \le X \le 3, 3 \le Y \le 4) = \int_0^3 \int_3^4 f(x,y)dxdy = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{\underline{r}(\theta)}^{\bar{r}(\theta)} f(r\cos\theta, r\sin\theta) r dr d\theta,$$

where $\underline{r}(\theta)$ and $\bar{r}(\theta)$ are shortest and longest distance between the origin and the points in $\{(x,y) : 0 \le x \le 3, 3 \le y \le 4, y = (\tan\theta)x\}$ (see Figure 7 for an illustration). It's easy to see we have the

bound $3 \le \underline{r}(\theta) \le \bar{r}(\theta) \le 5$ for any $\theta \in [\pi/4, \pi/2]$. Therefore, the objective function can be bounded by

$$\begin{aligned}
&\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{\underline{r}(\theta)}^{\bar{r}(\theta)} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta \\
\le &\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{3}^{5} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta \\
\le &\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \max_{3\le r\le 5} f(r\cos\theta, r\sin\theta) \int_{3}^{5} r\,dr\,d\theta \\
= &8\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} f(3\cos\theta, 3\sin\theta)\,d\theta,
\end{aligned}$$

where the last equality follows from $\max_{3\le r\le 5} f(r\cos\theta, r\sin\theta) = f(3\cos\theta, 3\sin\theta)$ by the definition of star-unimodality. On the other hand, by a similar argument, we have

$$\begin{aligned}
&\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{0}^{3} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta \\
\ge &\frac{9}{2}\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} f(3\cos\theta, 3\sin\theta)\,d\theta \\
\ge &\frac{9}{16}\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{\underline{r}(\theta)}^{\bar{r}(\theta)} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta.
\end{aligned}$$

where the last inequality is from the above bound. Then, the feasibility of f implies

$$\begin{aligned}
\frac{1}{4} &= \int_0^\infty \int_0^\infty f(x,y)\,dx\,dy \\
&\ge \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_0^3 f(r\cos\theta, r\sin\theta) r\,dr\,d\theta + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{\underline{r}(\theta)}^{\bar{r}(\theta)} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta \\
&\ge \frac{25}{16}\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{\underline{r}(\theta)}^{\bar{r}(\theta)} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta,
\end{aligned}$$

which implies that

$$\begin{aligned}
&P_f(0 \le X \le 3, 3 \le Y \le 4)\\
=&\int_0^3 \int_3^4 f(x,y)dxdy\\
=&\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_{\underline{r}(\theta)}^{\bar{r}(\theta)} f(r\cos\theta, r\sin\theta) r dr d\theta\\
\le&\frac{4}{25}.
\end{aligned}$$

Next, we construct the desired density f_δ for any $\delta > 0$. We use Figure 7 to help illustrate our construction. We construct two rays coming from $(-\delta, 0)$ with one passing $(0,3)$ and the other passing $(0,4)$. By a simply geometric argument, we can see the area of the orange triangle I is three times as large as the area of the blue triangle II for $\delta \in (0,9]$. Let f_δ be the density of the uniform distribution on the two triangles. Clearly, it satisfies the star unimodal constraint (8). Moreover, the area relation between triangles I and II ensures that

$$P_{f_\delta}(X \ge 0, Y \ge 0) = P_{f_\delta}(0 \le X \le 3, 3 \le Y \le 4) = 1/4.$$

So f_δ is the desired density. □

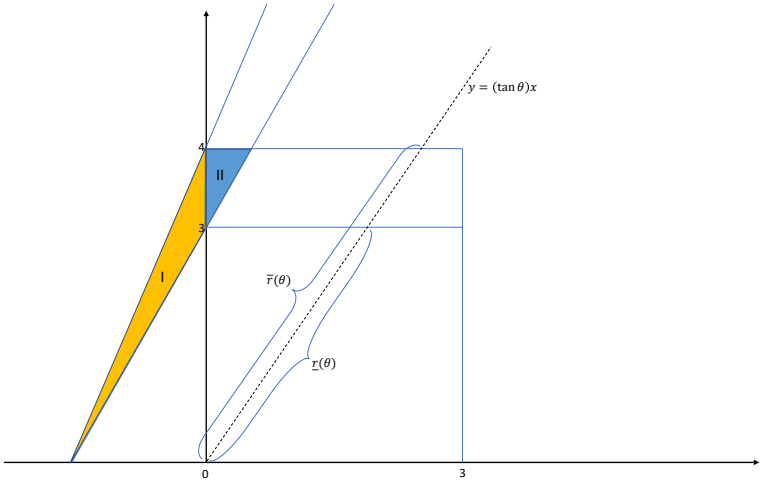


FIGURE 7. Illustration of the construction idea of f_δ

Proof of Proposition 2. Suppose that $x_0 = (x_{10}, x_{20}) \in \mathbb{R}^2_+$ belongs to the diagonal, i.e., $x_{10} = x_{20}$. To show f is α-unimodal about x_0 on $[x_{10}, \infty) \times [x_{20}, \infty)$, we need to show $t^{2-\alpha} f(tx + x_0) = t^{2-\alpha} f(tx_1 + x_{10}, tx_2 + x_{20})$ is non-increasing in $t \in (0, \infty)$ for any nonzero $x \in \mathbb{R}^2_+$. In fact, since $x_{10} = x_{20}$ and $x \in \mathbb{R}^2_+$, the following terms are both non-decreasing in t

$$\max\left(\arctan\left(\frac{tx_2 + x_{20}}{tx_1 + x_{10}}\right), \arctan\left(\frac{tx_1 + x_{10}}{tx_2 + x_{20}}\right)\right), \quad tx_1 + x_{10} + tx_2 + x_{20}.$$

Therefore, the density $f(tx + x_0)$ is non-increasing in t. Since $t^{2-\alpha}$ is also non-increasing in t, we obtain that $t^{2-\alpha} f(tx + x_0)$ is non-increasing in t, i.e., f is α-unimodal about x_0 on $[x_{10}, \infty) \times [x_{20}, \infty)$.

Now we assume $x_0 = (x_{10}, x_{20}) \in \mathbb{R}^2_+$ and $|x_{20} - x_{10}| > 4(\alpha - 2)/(4 - \pi)$. Since the density is symmetric about x_1 and x_2, we assume without loss of generality that $x_{20} - x_{10} > 4(\alpha - 2)/(4 - \pi)$. We consider the direction $(1, 0)$ and show $t^{2-\alpha} f(t + x_{10}, x_{20})$ can be strictly increasing for t in some interval. When $t \le x_{20} - x_{10}$, the logarithm of $t^{2-\alpha} f(t + x_{10}, x_{20})$ (discarding the constant C) is

$$g(t) = -\arctan\left(\frac{x_{20}}{t + x_{10}}\right)(t + x_{10} + x_{20}) + (2 - \alpha)\log t.$$

We will show $g'(t)$ is strictly positive for $t = x_{20} - x_{10}$ and consequently obtain that $g'(t)$ is strictly positive on $(x_{20} - x_{10} - \epsilon, x_{20} - x_{10}]$ for some small $\epsilon > 0$ due to the continuity of $g'(t)$, which will imply that $t^{2-\alpha} f(t + x_{10}, x_{20})$ is strictly increasing for $t \in (x_{20} - x_{10} - \epsilon, x_{20} - x_{10}]$. By direct computation, we have

$$g'(t) = \frac{x_{20}}{(t + x_{10})^2 + x_{20}^2}(t + x_{10} + x_{20}) - \arctan\left(\frac{x_{20}}{t + x_{10}}\right) + \frac{2 - \alpha}{t}.$$

When $t = x_{20} - x_{10}$, we have

$$g'(x_{20} - x_{10}) = 1 - \frac{\pi}{4} + \frac{2 - \alpha}{x_{20} - x_{10}} = \frac{1}{x_{20} - x_{10}}\left((x_{20} - x_{10})\left(1 - \frac{\pi}{4}\right) + 2 - \alpha\right) > 0$$

due to $x_{20} - x_{10} > 4(\alpha - 2)/(4 - \pi)$. This completes our proof. □

Proof of Lemma 2. Let us first show $\bar{K}$ and $\bar{K}^\circ$ are OU sets about x_0. In order to show $\bar{K}$ is OU, it suffices to show that if x is a limit point of K, then $x' \in \bar{K}$ if $x_0 \le x' \le x$. Suppose that the sequence $\{x_j, j \ge 1\} \subset K$ converges to x. Consider the sequence $\{\min(x', x_j), j \ge 1\}$ where the minimum is taken component-wise. By the definition of OU sets, this sequence must be in K. Besides, we notice that the limit point of this sequence is $\min(x', x) = x'$. Thus $x' \in \bar{K}$, i.e., $\bar{K}$ is an OU set about x_0. Now let's show $\bar{K}^\circ$ is also OU. Suppose $x \in \bar{K}^\circ$, i.e., there exists a sequence $\{x_j\} \subset K^\circ$ s.t. $x_j \to x$. We need to show $x' \in \bar{K}^\circ$ if $x_0 \le x' \le x$. By the definition of OU, $x_j \in K^\circ$

implies $x_0 < x_j$ and $x'' \in K^\circ$ for $x_0 < x'' \le x_j$. Therefore, if $x_0 < x' \le x$, we have $\min(x_j, x') \in K^\circ$. Since $\min(x_j, x') \to x'$, we know $x' \in \bar{K}^\circ$ for all x' satisfying $x_0 < x' \le x$. Moreover, since $\bar{K}^\circ$ is a closed set, we can get $x' \in \bar{K}^\circ$ for $x_0 \le x' \le x$, which means $\bar{K}^\circ$ is OU about x_0.

Then let us consider other claims in the lemma. Since the Lebesgue measure is complete, in order to show K is Lebesgue measurable and $\lambda(K^\circ) = \lambda(K) = \lambda(\bar{K})$, it suffices to show that $\lambda(\partial K) = 0$. This is already proven in [66], which concludes our proof. □

Proof of Theorem 1. Suppose there is a density $f(x)$ of P such that for every $s > 0$, the set

$$C_s = \{x \in \mathcal{D}_0 : f(x) \ge s\}$$

is OU about x_0. Write $W_{\bar{C}_s}$ as the uniform distribution on $\bar{C}_s$ and let $g(s) = \lambda(\bar{C}_s) \ge 0$. Notice that

$$f(x) = \int_0^{f(x)} 1 ds = \int_0^\infty I(x \in C_s) ds.$$

Thus for any measurable set B,

$$P(B) = \int_B f(x) dx = \int_B \int_0^\infty I(x \in C_s) ds dx = \int_0^\infty \lambda(B \cap C_s) ds.$$

By Lemma 2, we know that $\lambda(\partial C_s) = 0$. Thus, $\lambda(B \cap C_s) = \lambda(B \cap \bar{C}_s)$, which means

$$P(B) = \int_0^\infty \lambda(B \cap \bar{C}_s) ds = \int_0^\infty W_{\bar{C}_s}(B) g(s) ds.$$

Setting $B = \mathcal{D}_0$, we can see $P(B) = 1 = \int_0^\infty g(s) ds$ and thus $g(s)$ is a probability density on $(0, \infty)$. This proves the correctness of (12). By Lemma 2, we know that $\bar{C}_s$ is an OU set. So $W_{\bar{C}_s}$ is the uniform distribution on the OU set $\bar{C}_s$. Since the probability measure with the density $g(s)$ must be the limit of a sequence of discrete probability measures with finite support, we can see P is in the closed convex hull of the set of all uniform distributions on OU subsets of $\mathcal{D}_0$, i.e., P is an OU distribution.

To see the converse, we assume P is OU about x_0. Suppose that Q is the uniform distribution on an OU set $K \subset \mathcal{D}_0$ and $0 < \lambda(K) < \infty$. For a point $x \in \mathbb{R}^d$ and $\delta > 0$, we define the neighborhood $N_\delta(x) = \{y \in \mathbb{R}^d : \max_{1 \le i \le d} |x_i - y_i| < \delta\}$. For ease of illustration, we assume x_0 is the origin. Otherwise, we can perform a translation to achieve it. Then for $x, x' \in \mathcal{D}_0^\circ$ with $x \ge x'$ and $0 < \delta < \min_{1 \le i \le d} x_i'$, we have

$$y \in K \cap N_\delta(x) \Rightarrow y + x' - x \in K \cap N_\delta(x').$$

Since Lebesgue measure is translation invariant, we can get

$$\lambda(K \cap N_\delta(x')) \geq \lambda(K \cap N_\delta(x)).$$

Dividing by $\lambda(K)$, we have

$$Q(N_\delta(x')) \geq Q(N_\delta(x)). \tag{51}$$

Clearly, this relation also holds under the convex combinations of the uniform distributions on OU sets. Since P is OU about x_0, by definition, there is a sequence $\{Q_m, m \geq 1\}$ such that Q_m converges weakly to P, where Q_m's are the convex combinations of the uniform distributions on OU sets. Therefore, these Q_m's satisfy (51). Since P has a density, say f_0, we have

$$P(\partial N_\delta(x')) = P(\partial N_\delta(x)) = 0.$$

Weak convergence and (51) imply that

$$P(N_\delta(x')) \geq P(N_\delta(x)). \tag{52}$$

For $x \in \mathcal{D}_0^\circ$, we define

$$f(x) = \limsup_{\delta \downarrow 0} \frac{P(N_\delta(x))}{\lambda(N_\delta(x))} = \limsup_{\delta \downarrow 0} \frac{P(N_\delta(x))}{(2\delta)^d}.$$

By the Lebesgue differentiation theorem, $f(x) = f_0(x)$ a.e., which means $f(x)$ is also a density for P. Besides, (52) implies $f(x') \geq f(x)$ for $x, x' \in \mathcal{D}_0^\circ$ with $x \geq x'$. For $x \in \partial\mathcal{D}_0$, we simply define

$$f(x) = \sup_{y \in \mathcal{D}_0^\circ} f(y).$$

Then the density $f(x)$ is what we want.

Finally, note that only $W_{\bar{C}_s}$ with $\lambda(\bar{C}_s) > 0$ contributes to (12). Thus, to show (12) is indeed a Choquet representation, it suffices to show W_K is an extreme point in the class of OU distributions on $\mathcal{D}_0$ if K is an OU set about x_0 with $\lambda(K) > 0$. This is proved in Lemma 3. □

Proof of Lemma 3. Suppose W_K is not an extreme point, i.e., there exist two different OU distributions P_1, P_2 and a number $\eta \in (0,1)$ s.t.

$$W_K = \eta P_1 + (1-\eta) P_2. \tag{53}$$

By Lemma 2, we know W_K and W_{K° are the same distribution. Therefore, the support of W_K is the OU set $\bar{K}^\circ$ (closure of K°). Let K_1 and K_2 be the support of P_1 and P_2 respectively. We have

$\bar{K}^\circ = K_1 \cup K_2$. Since W_K is absolutely continuous with respect to Lebesgue measure, P_1 and P_2 are also absolutely continuous with respect to Lebesgue measure. Let f_{W_K}, f_1 and f_2 be densities of W_K, P_1 and P_2 respectively. By Theorem 1, we can without loss of generality assume that $f_1(x)$ and $f_2(x)$ are non-increasing in each component of x on $\mathcal{D}_0$, which also implies K_1 and K_2 are OU sets about x_0. Then (53) can be written as

$$f_{W_K} = \eta f_1 + (1-\eta) f_2, \text{ a.e..}$$

However, since f_{W_K} is constant on the OU set $\bar{K}^\circ$ and $f_1(x)$, $f_2(x)$ are non-increasing in each component of x, $f_1(x)$ and $f_2(x)$ must also be constant a.e. on $\bar{K}^\circ$. Thus, P_1 and P_2 are uniform distributions on K_1 and K_2 respectively. Since $\bar{K}^\circ$, K_1 and K_2 are all OU sets with positive Lebesgue measure, there exists an open set (close to x_0) $K' \subset \bar{K}^\circ \cap K_1 \cap K_2$. By (53), we have

$$\frac{\lambda(K')}{\lambda(\bar{K}^\circ)} = \eta \frac{\lambda(K')}{\lambda(K_1)} + (1-\eta)\frac{\lambda(K')}{\lambda(K_2)},$$

which implies $\lambda(\bar{K}^\circ) = \lambda(K_1) = \lambda(K_2)$. So we know W_K, P_1 and P_2 are the same distribution, which contradicts our assumption that P_1 and P_2 are different distributions. Therefore, W_K is an extreme point in the class of OU distributions about x_0. □

Proof of Proposition 3. Note that a closed rectangle in $\mathcal{D}_0$ which contains x_0 and has edges parallel to the coordinate axes is OU about x_0. Moreover, an OU set about x_0 is also star-shaped about x_0. Therefore, by the definitions of these three unimodalities, we have

$$\begin{aligned}
&\{\text{distributions on } \mathcal{D}_0 \text{ which is block unimodal about } x_0\}\\
\subset&\{\text{distributions on } \mathcal{D}_0 \text{ which is OU about } x_0\}\\
\subset&\{\text{distributions on } \mathcal{D}_0 \text{ which is star unimodal about } x_0\}.
\end{aligned}$$

It remains to prove that the inclusions above are proper. For simplicity, we only show it in the bivariate case. Higher dimensional cases give us no additional insight but only the complication of notations. Without loss of generality, we assume x_0 is the origin.

To show the first inclusion is proper, consider a probability distribution P with the following density:

$$f(x,y) = \begin{cases} 1/7 & \text{if } (x,y) \in (1,2] \times (1,2] \\ 2/7 & \text{if } (x,y) \in \{[0,2] \times [0,2]\} \backslash \{(1,2] \times (1,2]\} \\ 0 & \text{otherwise.} \end{cases}$$

Let P_1 be the uniform distribution on $[0,2]\times[0,2]$ and P_2 be the uniform distribution on $\{[0,2]\times[0,2]\}\backslash\{(1,2]\times(1,2]\}$. Then P is a convex combination of P_1 and P_2 ($P=4P_1/7+3P_2/7$). Note that both sets are OU about the origin. Therefore, P is OU about the origin by definition. However, P is not block unimodal about the origin. To see this, consider four points $(1/2,1/2)$, $(3/2,3/2)$, $(3/2,1/2)$ and $(1/2,3/2)$ at which $f(x,y)$ is continuous. If P is block unimodal, we must have

$$f\left(\frac{1}{2},\frac{1}{2}\right)+f\left(\frac{3}{2},\frac{3}{2}\right)\geq f\left(\frac{3}{2},\frac{1}{2}\right)+f\left(\frac{1}{2},\frac{3}{2}\right)$$

according to Section 2.2 in [30] (or we can show this by the similar proof of the "only if" part of Theorem 1). However,

$$f\left(\frac{1}{2},\frac{1}{2}\right)+f\left(\frac{3}{2},\frac{3}{2}\right)=\frac{3}{7}<\frac{4}{7}=f\left(\frac{3}{2},\frac{1}{2}\right)+f\left(\frac{1}{2},\frac{3}{2}\right),$$

which means P is not block unimodal about the origin. This proves the first inclusion is proper.

To show the second inclusion is also proper, we consider the density mentioned in Section 3:

$$f(x,y)=C\exp\left(-\max\left(\arctan\left(\frac{y}{x}\right),\arctan\left(\frac{x}{y}\right)\right)(x+y)\right),x\geq 0,y\geq 0,$$

where C is the normalizing constant. For any ray $\{(r\cos\theta,r\sin\theta):r\geq 0\}$, $\theta\in[0,\pi/2]$, we can see $f(x,y)$ is non-increasing along it. So according to Section 2.2 in [30], $f(x,y)$ is star unimodal about the origin. However, it's not OU about the origin because by direct calculation we have

$$f(1,2)\approx 0.036C<0.043C\approx f(2,2),$$

which violates Theorem 1. Thus the second inclusion is also proper. □

Proof of Corollary 1. This follows directly from Lemma 1. □

Proof of Proposition 4. The proof is constructive. Suppose we remove the moment constraint (16) from (20) and get the following problem

$$\begin{aligned}
&\max\ P((X_1,\ldots,X_d)\in S)\\
&\text{subject to } l_{\bar{F}}\leq\bar{F}(x_0)\leq u_{\bar{F}}\\
&\qquad f_{X_i}(x_{i0})\leq u_{X_i},i=1,\ldots,d\\
&\qquad f(x')\geq f(x)\text{ for }x\geq x'\geq x_0.
\end{aligned}\tag{54}$$

We can see there is a trivial bound for the optimal value of the problem (54), i.e.,

$$P((X_1,\dots,X_d)\in S)\le \bar{F}(x_0)\le u_{\bar{F}}. \tag{55}$$

We will construct a sequence of feasible distributions of $(X_1,\dots,X_d)$ s.t.

$$P((X_1,\dots,X_d)\in S)\to u_{\bar{F}}, \tag{56}$$

which implies the optimal value of (54) is exactly $u_{\bar{F}}$.

Consider a sequence of rectangles $\{R_i, i\ge 1\}$ in $\mathcal{D}_0$ defined as

$$R_i=\{(x_1,\dots,x_d): x_{j0}\le x_j\le x_{j0}+m, j=1,\dots,d-1 \text{ and } x_{d0}\le x_d\le x_{d0}+i\},$$

where $1/m<\min(u_{X_1},\dots,u_{X_d})$ is a fixed number. Correspondingly, we define F_i as the uniform distribution on R_i with total mass $u_{\bar{F}}$. The density f_i of the distribution F_i is given by

$$f_i(x_1,\dots,x_d)=\frac{u_{\bar{F}}}{m^{d-1}i}I(x_{j0}\le x_j\le x_{j0}+m, j=1,\dots,d-1)I(x_{d0}\le x_d\le x_{d0}+i).$$

Thus, the constraint that $f(x')\ge f(x)$ for $x\ge x'\ge x_0$ is satisfied. Besides, the truncated marginal density f_{X_j} at point x_{j0} satisfies

$$f_{X_j}(x_{j0})=\frac{u_{\bar{F}}}{m}\le\frac{1}{m}<u_{X_j} \text{ for } j=1,\dots,d-1,$$

and

$$f_{X_d}(x_{d0})=\frac{u_{\bar{F}}}{i}\le\frac{1}{i}<u_{X_d} \text{ when } i \text{ is large enough.}$$

So when i is large enough, the distribution F_i is feasible to the problem (54).

Now let us consider the objective value of F_i. Recall that $S=\{(x_1,\dots,x_d)\in\mathcal{D}_0: x_d\ge g(x_1,\dots,x_{d-1})\}$ and $g(x_1,\dots,x_{d-1})\ge x_{d0}$.

$$\begin{aligned}&P_{F_i}((X_1,\dots,X_d)\in S)\\ &=\int\frac{u_{\bar{F}}}{m^{d-1}i}I(x_j\in[x_{j0},x_{j0}+m], j=1,\dots,d-1, x_d\in[g(x_1,\dots,x_{d-1}),x_{d0}+i])dx_1\cdots dx_d.\end{aligned}$$

According to our assumption, g is bounded on compact sets. So $\exists\, M \geq x_{d0}$ s.t. $g(x_1,\ldots,x_{d-1}) \leq M$ when $x_{j0} \leq x_j \leq x_{j0}+m$, $j=1,\ldots,d-1$. Then when $i > M - x_{d0}$, we have

$$\begin{aligned}
&P_{F_i}((X_1,\ldots,X_d)\in S)\\
\geq &\int \frac{u_{\bar{F}}}{m^{d-1}i} I(x_{j0}\leq x_j \leq x_{j0}+m, j=1,\ldots,d-1) I(M \leq x_d \leq x_{d0}+i) dx_1\cdots dx_d\\
=&u_{\bar{F}}\frac{i+x_{d0}-M}{i} \to u_{\bar{F}},
\end{aligned}$$

i.e., a sequence of feasible distributions satisfying (56) exists.

Finally, it follows by the trivial bound (55) that the optimal value of the problem (54) is exactly $u_{\bar{F}}$. □

Proof of Lemma 4. Suppose f is a feasible solution to the problem (22) and $C_s = \{x \in \mathcal{D}_0 : f(x) \geq s\}$. We will show the objective function and the constraints in the problem (23) are equivalent to those in the problem (22).

Notice that $\{C_s, s>0\}$ is a non-increasing sequence of OU sets by Theorem 1. First, we show two properties of $\{C_s, s>0\}$: $\lambda_d(C_s) < \infty$ for any $s>0$, and $\lambda_d(C_s)=0$ implies $C_t = \emptyset$ for $t > s$. The first one is obviously true otherwise $\bar{F}(x_0) \geq s\lambda_d(C_s) = \infty$. Now let us show the second property. Suppose that $\lambda_d(C_s)=0$ and $t>s$. If $C_t \neq \emptyset$, then $x_0 \in C_t$ since x_0 is in any nonempty OU set about x_0. By the definition $C_t = \{x \in \mathcal{D}_0 : f(x) \geq t\}$, we know that $f(x_0) \geq t > s$. Since the regularity constraint on f says

$$f(x_0) = \limsup_{y \downarrow x_0, y \in \mathcal{D}_0^\circ} f(y),$$

there exists $y_0 \in \mathcal{D}_0^\circ$ s.t. $f(y_0) > s$. It follows from the OU property that $f(x) > s$ for $x_0 \leq x \leq y_0$. Therefore, $\lambda_d(C_s) > 0$, which contradicts our assumption. So we must have $C_t = \emptyset$.

Define $s_0 = \sup\{s>0 : \lambda(C_s) > 0\}$. From the last paragraph, we know that $0 < \lambda_d(C_s) < \infty$ for $0 < s < s_0$ and $C_s = \emptyset$ for $s > s_0$. For the probability measure P induced by the density f, from the proof of Theorem 1, we know that for any measurable set B, the following representation holds

$$P(B) = \int_0^\infty \lambda_d(B\cap \bar{C}_s) ds = \int_0^{s_0} \lambda_d(B\cap \bar{C}_s) ds = \int_0^{s_0} \frac{\lambda_d(B\cap \bar{C}_s)}{\lambda_d(\bar{C}_s)} \lambda_d(\bar{C}_s) ds,$$

where the last equality holds because $0 < \lambda_d(C_s) < \infty$ for $0 < s < s_0$ and $\lambda_d(C_s) = \lambda_d(\bar{C}_s)$ by Lemma 2. When $s > s_0$, $C_s = \bar{C}_s = \emptyset$ and thus $\lambda_d(B\cap\bar{C}_s)/\lambda_d(\bar{C}_s) = 0$ by our definition. Then we can extend the integral interval from $(0, s_0)$ to $(0,\infty)$ and get

$$P(B) = c\int_0^\infty \frac{\lambda_d(B\cap\bar{C}_s)}{\lambda_d(\bar{C}_s)} g(s) ds, \tag{57}$$

where $g(s) = \lambda_d(\bar{C}_s)/c$. In fact, $g(s)$ is a probability density on $(0,\infty)$, which can be verified by setting $B = \mathcal{D}_0$ in (57) and $P(\mathcal{D}_0) = c$. The representation (57) ensures the correctness of the representations of the objective function and the last constraint in the problem (23).

Now let us show the correctness of the representation of the first constraint in the problem (23). Take $i = 1$ as an example. Notice that

$$f(x_{10}, x_2, \ldots, x_d) = \int_0^{f(x_{10},x_2,\ldots,x_d)} 1 ds = \int_0^{\infty} I((x_{10}, x_2, \ldots, x_d) \in C_s) ds.$$

Since $C_s = \emptyset$ for $s > s_0$, it can also be written as

$$f(x_{10}, x_2, \ldots, x_d) = \int_0^{s_0} I((x_{10}, x_2, \ldots, x_d) \in C_s) ds.$$

Thus, the marginal density $f_{X_1}(x_{10})$ is given by

$$\begin{aligned}
f_{X_1}(x_{10}) &= \int_{x_{20}}^{\infty} \cdots \int_{x_{d0}}^{\infty} f(x_{10}, x_2, \ldots, x_d) dx_2 \cdots dx_d \\
&= \int_{x_{20}}^{\infty} \cdots \int_{x_{d0}}^{\infty} \int_0^{s_0} I((x_{10}, x_2, \ldots, x_d) \in C_s) ds dx_2 \cdots dx_d \\
&= \int_0^{s_0} \int_{x_{20}}^{\infty} \cdots \int_{x_{d0}}^{\infty} I((x_{10}, x_2, \ldots, x_d) \in C_s) dx_2 \cdots dx_d ds \\
&= \int_0^{s_0} \lambda_{d-1}(C_{s,1}) ds,
\end{aligned}$$

where $C_{s,1} = \{(x_2, \ldots, x_d) : (x_{10}, x_2, \ldots, x_d) \in C_s\}$ is slice of C_s on the plane $x = x_{10}$. Next, let's show $\lambda_{d-1}(C_{s,1}) = \lambda_{d-1}(\bar{C}_{s,1})$, where $\bar{C}_{s,1}$ is the slice of $\bar{C}_s$ on the plane $x = x_{10}$. We will show $\bar{C}_{s,1} \subset \overline{C_{s,1}}$, where $\overline{C_{s,1}}$ is the closure of $C_{s,1}$. Suppose $(x_2, \ldots, x_d) \in \bar{C}_{s,1}$, i.e., $(x_{10}, x_2, \ldots, x_d) \in \bar{C}_s$. Then there is a sequence $\{(x_{1k}, x_{2k}, \ldots, x_{dk}), k \geq 1\} \subset C_s$ such that

$$(x_{1k}, x_{2k}, \ldots, x_{dk}) \to (x_{10}, x_2, \ldots, x_d).$$

By the OU property, $\{(x_{10}, x_{2k}, \ldots, x_{dk}), k \geq 1\}$ are also in C_s, which implies $\{(x_{2k}, \ldots, x_{dk}), k \geq 1\} \subset C_{s,1}$. Thus, its limiting point $(x_2, \ldots, x_d)$ is in $\overline{C_{s,1}}$. This means $\bar{C}_{s,1} \subset \overline{C_{s,1}}$. Clearly, $C_{s,1} \subset \bar{C}_{s,1}$ and $C_{s,1}$ is OU about $(x_{20}, \ldots, x_{d0})$ in $\mathbb{R}^{d-1}$. By Lemma 2, we can get

$$\lambda_{d-1}(C_{s,1}) \leq \lambda_{d-1}(\bar{C}_{s,1}) \leq \lambda_{d-1}(\overline{C_{s,1}}) = \lambda_{d-1}(C_{s,1}) \Rightarrow \lambda_{d-1}(C_{s,1}) = \lambda_{d-1}(\bar{C}_{s,1}).$$

Thus, we can rewrite $f_{X_1}(x_{10})$ as

$$f_{X_1}(x_{10}) = \int_0^{s_0} \lambda_{d-1}(\bar{C}_{s,1})ds = \int_0^{s_0} \frac{\lambda_{d-1}(\bar{C}_{s,1})}{\lambda_d(\bar{C}_s)}\lambda_d(\bar{C}_s)ds = c\int_0^{s_0} \frac{\lambda_{d-1}(\bar{C}_{s,1})}{\lambda_d(\bar{C}_s)}g(s)ds,$$

where $g(s) = \lambda_d(\bar{C}_s)/c$ is defined above. When $s > s_0$, $C_s = \bar{C}_s = \emptyset$ and thus $\lambda_{d-1}(\bar{C}_{s,1})/\lambda_d(\bar{C}_s) = 0$ by our definition. So we can extend the integral interval from $(0, s_0)$ to $(0, \infty)$ and get

$$f_{X_1}(x_{10}) = c\int_0^{\infty} \frac{\lambda_{d-1}(\bar{C}_{s,1})}{\lambda_d(\bar{C}_s)}g(s)ds.$$

Thus, the representation of the first constraint of the problem (23) is correct. This concludes our proof of this lemma. □

Proof of Lemma 5. It suffices to show for any feasible solution to the problem (23), there exists a feasible solution to the problem (24) with the same objective value. Consider a density $f(x)$ on $\mathcal{D}_0$ with total mass c and satisfying the last two constraints in (22). Let $\{C_s, s > 0\}$ and $g(s)$ be the sequence of OU sets and the probability density defined in Lemma 4. In the proof of Lemma 4, we know there exists $s_0 \in (0, \infty)$ s.t. $0 < \lambda_d(C_s) = \lambda_d(\bar{C}_s) < \infty$ for $0 < s < s_0$ and $C_s = \emptyset$ for $s > s_0$. Therefore, $g(s) = 0$ for $s > s_0$ by its definition. Consider the slice of $\bar{C}_s$, e.g., $\bar{C}_{s,1}$. By the property of OU sets, we know $\bar{C}_s \subset \mathbb{R} \times \bar{C}_{s,1}$. Therefore, $\lambda_d(\bar{C}_s) > 0$ implies $\lambda_{d-1}(\bar{C}_{s,1}) > 0$ for $0 < s < s_0$. Similarly, we can also get $\lambda_{d-1}(\bar{C}_{s,i}) > 0$ for $0 < s < s_0$ and $i = 1, \dots, d$. Moreover, we must have $\lambda_{d-1}(\bar{C}_{s,i}) < \infty$ for any $s > 0$ and $i = 1, \dots, d$. If not, suppose without loss of generality that $\lambda_{d-1}(\bar{C}_{s_1,1}) = \infty$ for some $s_1 > 0$. Since $\bar{C}_{s,1}$ is non-increasing in s, we have $\lambda_{d-1}(\bar{C}_{s,1}) = \infty$ for $0 < s \le s_1$. Now we consider the first constraint in (23) with $i = 1$:

$$\begin{aligned}
&\frac{u_{X_1}}{c} \\
\ge &\int_0^{\infty} \frac{\lambda_{d-1}(\bar{C}_{s,1})}{\lambda_d(\bar{C}_s)}g(s)ds \\
\ge &\int_0^{\min(s_0,s_1)} \frac{\lambda_{d-1}(\bar{C}_{s,1})}{\lambda_d(\bar{C}_s)}g(s)ds \\
= &\frac{1}{c}\int_0^{\min(s_0,s_1)} \lambda_{d-1}(\bar{C}_{s,1})ds \\
= &\infty,
\end{aligned}$$

which contradicts to the assumption that f is a feasible solution to the problem (22). In summary, we know that $\lambda_d(\bar{C}_s) \in (0, \infty), \lambda_{d-1}(\bar{C}_{s,i}) \in (0, \infty)$ for any $s \in (0, s_0)$ and $i = 1, \dots, d$, and $g(s) = 0$ for $s > s_0$.

Now we construct a feasible solution to the problem (24) with the same objective value as $f(x)$. Notice that $s \mapsto \tan(\pi s/2s_0)$ is a bijection from $(0, s_0)$ into $(0, \infty)$ and

$$\int_0^{s_0} \varphi(s)g(s)ds = \int_0^{\infty} \varphi\left(\frac{2s_0 \arctan(u)}{\pi}\right) g\left(\frac{2s_0 \arctan(u)}{\pi}\right) \frac{2s_0}{\pi(1+u^2)} du \tag{58}$$

for any non-negative measurable function φ by the change of variable $u = \tan(\pi s/2s_0)$. Therefore, if we define R_s by

$$R_s = \bar{C}_{\frac{2s_0 \arctan(u)}{\pi}}, s \in (0, \infty)$$

and define $G(s)$ as the probability distribution with density

$$G'(s) = g\left(\frac{2s_0 \arctan(s)}{\pi}\right) \frac{2s_0}{\pi(1+s^2)}, s \in (0, \infty),$$

it follows by (58) that the objective value and the constraints in (23) and (24) are exactly the same. Further, R_s satisfies $\lambda_d(R_s) \in (0, \infty), \lambda_{d-1}(R_{s,i}) \in (0, \infty)$ for any $i = 1, \dots, d$ and $s > 0$, which means $\{R_s, s > 0\}$ and $G(s)$ are feasible to the problem (24). This concludes our proof. □

Proof of Corollary 2. This follows directly from Lemma 5 when $d = 2$. □

Proof of Lemma 6. By the definition of R in (28) and the constraints of h in (27), we can easily see that R is a closed OU set with $R^X \le R_0^X$ and $R^Y = h(x_0) \le h_0(x_0+) \le R_0^Y$. By the first constraint in (27), we have

$$\begin{aligned}
\lambda(R) &= \int_{x_0}^{x_0+r_X} (h(x) - y_0)dx = \sum_{i=0}^{n_{R_0}} \int_{x_i}^{x_{i+1}} (h(x) - y_0)dx \\
&= \sum_{i=0}^{n_{R_0}} \lambda(\{(x, y) : x_i \le x \le x_{i+1}, y \ge y_0\} \cap R_0) \\
&= \lambda(\{(x, y) : x_0 \le x \le x_0 + r_X, y \ge y_0\} \cap R_0) = \lambda(R_0).
\end{aligned}$$

So it remains to verify $\lambda(\{(x, y) : x_{1i} \le x \le x_{2i}, y_{1i} \le y \le y_{2i}\} \cap R) = \lambda(\{(x, y) : x_{1i} \le x \le x_{2i}, y_{1i} \le y \le y_{2i}\} \cap R_0), i = 1, \dots, n$. There are several cases for these equalities, considering whether $h_0^{-1}(y_{1i})$ and $h_0^{-1}(y_{2i})$ are well defined and all possible orders of $x_{1i}, x_{2i}, h_0^{-1}(y_{1i})$ and $h_0^{-1}(y_{2i})$. Here we only consider one case and other cases can be proved similarly. Suppose that both $h_0^{-1}(y_{1i})$ and $h_0^{-1}(y_{2i})$ are well defined and $x_{1i} < h_0^{-1}(y_{2i}) < h_0^{-1}(y_{1i}) < x_{2i}$. In this case, $x_{1i}, h_0^{-1}(y_{2i}), h_0^{-1}(y_{1i})$ and x_{2i} are distinct in the ordered sequence $x_0 < x_1 < \cdots < x_{n_{R_0}+1}$. By the definition $h_0^{-1}(y) = \sup\{x : h_0(x) \ge y\}$

and the left-continuity of h_0, we must have $h_0(h_0^{-1}(y)) \geq y$ and $h_0(h_0^{-1}(y)+) \leq y$. Notice that we have

$$\begin{aligned}
&\lambda(\{(x,y) : x_{1i} \leq x \leq x_{2i}, y_{1i} \leq y \leq y_{2i}\} \cap R_0) \\
=&(y_{2i} - y_{1i})(h_0^{-1}(y_{2i}) - x_{1i}) + \int_{h_0^{-1}(y_{2i})}^{h_0^{-1}(y_{1i})} (h_0(x) - y_{1i})dx \\
=&(y_{2i} - y_{1i})(h_0^{-1}(y_{2i}) - x_{1i}) + \lambda(\{(x,y) : h_0^{-1}(y_{2i}) \leq x \leq h_0^{-1}(y_{1i}), y \geq y_0\} \cap R_0) \\
&- (y_{1i} - y_0)(h_0^{-1}(y_{1i}) - h_0^{-1}(y_{2i})).
\end{aligned}$$

To see how the first equality is obtained, recall that $R_0 = \{(x,y) : y_0 \leq y \leq h_0(x), x_0 \leq x \leq x_0 + R_0^X\}$. Therefore, the intersection is essentially $\{(x,y) : x_{1i} \leq x \leq x_{2i}, \min(y_{1i}, h_0(x)) \leq y \leq \min(y_{2i}, h_0(x))\}$. Since we assume $x_{1i} < h_0^{-1}(y_{2i}) < h_0^{-1}(y_{1i}) < x_{2i}$ in this case and h_0 is non-decreasing, the boundary of y reduces to $y_{1i} \leq y \leq y_{2i}$ when $x \in [x_{1i}, h_0^{-1}(y_{2i})]$, reduces to $y_{1i} \leq y \leq h_0(x)$ when $x \in (h_0^{-1}(y_{2i}), h_0^{-1}(y_{1i})]$ and reduces to $h_0(x) \leq y \leq h_0(x)$ when $x \in (h_0^{-1}(y_{1i}), x_{2i}]$. Now let us calculate $\lambda(\{(x,y) : x_{1i} \leq x \leq x_{2i}, y_{1i} \leq y \leq y_{2i}\} \cap R)$. By the second and third constraints in (27), we can see: $h(x) \geq h_0(h_0^{-1}(y_{2i})) \geq y_{2i}$ if $x \in (x_{1i}, h_0^{-1}(y_{2i})]$, $y_{1i} \leq h_0(h_0^{-1}(y_{1i})) \leq h(x) \leq h_0(h_0^{-1}(y_{2i})+) \leq y_{2i}$ if $x \in (h_0^{-1}(y_{2i}), h_0^{-1}(y_{1i})]$ and $h(x) \leq h_0(h_0^{-1}(y_{1i})+) \leq y_{1i}$ if $x \in (h_0^{-1}(y_{1i}), x_{2i}]$. So we have

$$\begin{aligned}
&\lambda(\{(x,y) : x_{1i} \leq x \leq x_{2i}, y_{1i} \leq y \leq y_{2i}\} \cap R) \\
=&(y_{2i} - y_{1i})(h_0^{-1}(y_{2i}) - x_{1i}) + \int_{h_0^{-1}(y_{2i})}^{h_0^{-1}(y_{1i})} (h(x) - y_{1i})dx.
\end{aligned}$$

Note that

$$\begin{aligned}
&\int_{h_0^{-1}(y_{2i})}^{h_0^{-1}(y_{1i})} (h(x) - y_{1i})dx \\
=&\int_{h_0^{-1}(y_{2i})}^{h_0^{-1}(y_{1i})} (h(x) - y_0)dx - (y_{1i} - y_0)(h_0^{-1}(y_{1i}) - h_0^{-1}(y_{2i})) \\
=&\lambda(\{(x,y) : h_0^{-1}(y_{2i}) \leq x \leq h_0^{-1}(y_{1i}), y \geq y_0\} \cap R_0) - (y_{1i} - y_0)(h_0^{-1}(y_{1i}) - h_0^{-1}(y_{2i})),
\end{aligned}$$

where the second equality follows from the first constraint in (27). It implies $\lambda(\{(x,y): x_{1i} \le x \le x_{2i}, y_{1i} \le y \le y_{2i}\} \cap R) = \lambda(\{(x,y): x_{1i} \le x \le x_{2i}, y_{1i} \le y \le y_{2i}\} \cap R_0)$ holds in this case. For other cases, the proofs are similar and thus are omitted. This concludes our proof. □

Proof of Lemma 7. Since the modified function h_0 (via redefining $h_0(x_0) = h_0(x_0+)$) is a feasible solution to (30), by (29) and the optimality of h^*, we have $\lambda(S \cap \tilde{R}) \ge \lambda(S \cap R_0)$. Moreover, by Lemma 6, $\tilde{R}$ satisfies all the remaining requirements of the dominating OU set of R_0. □

To prove Lemma 8, we first prove the following simple lemma.

LEMMA 12. *Suppose $a_3 > a_2 > a_1$ and c are some constants in $\mathbb{R}$, and $g : (a_1, a_3] \mapsto (-\infty, \infty]$. Consider the following function*

$$H(y) = \int_{a_1}^{a_2} (y - g(x))_+ \, dx + \int_{a_2}^{a_3} \left(\frac{c - (a_2 - a_1)\, y}{a_3 - a_2} - g(x) \right)_+ dx$$

whose domain is $y \in [\underline{y}, \bar{y}]$. Then the maximum value of $H(y)$ is attained at the endpoints of the domain.

Proof of Lemma 12. It suffices to show that $H(y)$ is a convex function. For $y_1, y_2 \in [\underline{y}, \bar{y}]$ and $\eta \in [0,1]$, we have

$$\begin{aligned}
&H(\eta y_1 + (1-\eta) y_2) \\
=&\int_{a_1}^{a_2} (\eta y_1 + (1-\eta) y_2 - g(x))_+ \, dx + \int_{a_2}^{a_3} \left(\frac{c - (a_2 - a_1)(\eta y_1 + (1-\eta) y_2)}{a_3 - a_2} - g(x) \right)_+ dx \\
=&\int_{a_1}^{a_2} (\eta (y_1 - g(x)) + (1-\eta)(y_2 - g(x)))_+ \, dx \\
&+ \int_{a_2}^{a_3} \left(\eta \left(\frac{c - (a_2 - a_1) y_1}{a_3 - a_2} - g(x) \right) + (1-\eta) \left(\frac{c - (a_2 - a_1) y_2}{a_3 - a_2} - g(x) \right) \right)_+ dx \\
\le&\int_{a_1}^{a_2} [\eta (y_1 - g(x))_+ + (1-\eta)(y_2 - g(x))_+] \, dx \\
&+ \int_{a_2}^{a_3} \left[\eta \left(\frac{c - (a_2 - a_1) y_1}{a_3 - a_2} - g(x) \right)_+ + (1-\eta) \left(\frac{c - (a_2 - a_1) y_2}{a_3 - a_2} - g(x) \right)_+ \right] dx \\
=&\eta H(y_1) + (1-\eta) H(y_2).
\end{aligned}$$

Therefore, $H(y)$ is a convex function. □

Now, let us show Lemma 8.

Proof of Lemma 8. We define a subproblem of the problem (33) by adding an additional constraint:

$$\begin{aligned} \max \quad & \int_{\underline{x}}^{\bar{x}} (h(x) - g(x))_+ dx \\ \text{subject to} \quad & \int_{\underline{x}}^{\bar{x}} h(x) dx = C \\ & \underline{b} \le h(x) \le \bar{b}, x \in (\underline{x}, \bar{x}] \\ & h \text{ is a non-increasing left-continuous step function,} \end{aligned} \tag{59}$$

where the step function is in the form of

$$h(x) = \begin{cases} y_1, & a_0 < x \le a_1 \\ y_2, & a_1 < x \le a_2 \\ \cdots & \\ y_n, & a_{n-1} < x \le a_n \end{cases},$$

where $n \in \mathbb{N}$, $\underline{x} = a_0 < a_1 < \cdots < a_n = \bar{x}$ and y_i's are some constants. To clarify, here we abuse the notations a_i and n. They are not the constants in the formulation of our DRO problem (13). We aim at showing the optimal value of the problem (59) is the same as that of the problem (33). It suffices to show that for any feasible solution h to the problem (33), there exists a sequence of feasible solutions $\{h_n, n \ge 1\}$ to the problem (59), s.t.

$$\int_{\underline{x}}^{\bar{x}} (h_n(x) - g(x))_+ dx \to \int_{\underline{x}}^{\bar{x}} (h(x) - g(x))_+ dx. \tag{60}$$

Now fix any feasible solution h to the problem (33). For $n \ge 1$, let's construct h_n. For simplicity, we extend the definition of h to $[\underline{x}, \bar{x}]$ by defining $h(\underline{x}) = h(\underline{x}+)$. For $i \in \{0, 1, \ldots, n-1\}$, since h is non-increasing, we have

$$h\left(\underline{x} + \frac{(i+1)(\bar{x}-\underline{x})}{n}\right)\frac{(\bar{x}-\underline{x})}{n} \le \int_{\underline{x}+\frac{i(\bar{x}-\underline{x})}{n}}^{\underline{x}+\frac{(i+1)(\bar{x}-\underline{x})}{n}} h(x)\, dx \le h\left(\underline{x} + \frac{i(\bar{x}-\underline{x})}{n}\right)\frac{(\bar{x}-\underline{x})}{n}.$$

Therefore, there is a real number $h_{i,n}$ satisfying

$$h_{i,n} \in \left[h\left(\underline{x} + \frac{(i+1)(\bar{x}-\underline{x})}{n}\right), h\left(\underline{x} + \frac{i(\bar{x}-\underline{x})}{n}\right)\right]$$

s.t.

$$\int_{\underline{x}+\frac{i(\bar{x}-\underline{x})}{n}}^{\underline{x}+\frac{(i+1)(\bar{x}-\underline{x})}{n}} h(x)\,dx = h_{i,n}\frac{(\bar{x}-\underline{x})}{n}.$$

Then our h_n is defined as

$$h_n(x) = h_{i,n}, \forall x \in \left(\underline{x}+\frac{i(\bar{x}-\underline{x})}{n}, \underline{x}+\frac{(i+1)(\bar{x}-\underline{x})}{n}\right].$$

We can see h_n is a feasible solution to the problem (59) by our construction. Let us verify (60).

$$\begin{aligned}
&\left|\int_{\underline{x}}^{\bar{x}} (h_n(x)-g(x))_+\,dx - \int_{\underline{x}}^{\bar{x}} (h(x)-g(x))_+\,dx\right| \\
\le &\int_{\underline{x}}^{\bar{x}} |\,(h_n(x)-g(x))_+ - (h(x)-g(x))_+\,|dx \\
\le &\int_{\underline{x}}^{\bar{x}} |\,(h_n(x)-g(x)) - (h(x)-g(x))\,|dx \\
= &\sum_{i=0}^{n-1} \int_{\underline{x}+\frac{i(\bar{x}-\underline{x})}{n}}^{\underline{x}+\frac{(i+1)(\bar{x}-\underline{x})}{n}} |h_n(x)-h(x)|\,dx \\
\le &\sum_{i=0}^{n-1} \int_{\underline{x}+\frac{i(\bar{x}-\underline{x})}{n}}^{\underline{x}+\frac{(i+1)(\bar{x}-\underline{x})}{n}} \left(h\left(\underline{x}+\frac{i(\bar{x}-\underline{x})}{n}\right) - h\left(\underline{x}+\frac{(i+1)(\bar{x}-\underline{x})}{n}\right)\right) dx \\
= &\frac{(\bar{x}-\underline{x})}{n}(h(\underline{x})-h(\bar{x})) \to 0
\end{aligned}$$

as $n \to \infty$. Thus, the optimal value of the problem (59) is the same as that of the problem (33).

Next, we define a subproblem of the problem (59) by adding an additional constraint again:

$$\begin{aligned}
\max \quad & \int_{\underline{x}}^{\bar{x}} (h(x)-g(x))_+ dx \\
\text{subject to} \quad & \int_{\underline{x}}^{\bar{x}} h(x)dx = C \\
& \underline{b} \le h(x) \le \bar{b}, x \in (\underline{x}, \bar{x}] \\
& h \text{ is a non-increasing left-continuous step function with at most three steps.}
\end{aligned} \tag{61}$$

Again, we aim at showing the optimal value of the problem (61) is the same as that of the problem (59). It suffices to show that for any feasible solution h_1 to the problem (59), there is a feasible solution h_2 to the problem (61) s.t.

$$\int_{\underline{x}}^{\bar{x}} (h_1(x) - g(x))_+ dx \le \int_{\underline{x}}^{\bar{x}} (h_2(x) - g(x))_+ dx. \tag{62}$$

Notice that for any feasible solution h_1 of the problem (59), it can be uniquely written as

$$h_1(x) = \begin{cases} y_1, & a_0 < x \le a_1 \\ y_2, & a_1 < x \le a_2 \\ \cdots & \\ y_n, & a_{n-1} < x \le a_n \end{cases},$$

where $\underline{b} \le y_n < y_{n-1} < \cdots < y_1 \le \bar{b}$ and $\underline{x} = a_0 < a_1 < \cdots < a_n = \bar{x}$. Now let's prove (62) by induction with respect to n (hereafter we will call n the number of steps of h_1). When $n \le 3$, $h_1(x)$ is also a feasible solution to the problem (61). So we can choose $h_2 = h_1$ and then (62) holds. Assume that when $n \le k$ ($k \ge 3$), the conclusion is true. Now suppose $n = k+1 \ge 4$.

Let's consider the steps y_{k-1} and y_k, i.e., consider the interval $(a_{k-2}, a_k]$. So the objective value in this interval is given by

$$\int_{a_{k-2}}^{a_{k-1}} (y_{k-1} - g(x))_+ dx + \int_{a_{k-1}}^{a_k} (y_k - g(x))_+ dx.$$

We want to change the function values in $(a_{k-2}, a_{k-1}]$ and $(a_{k-1}, a_k]$ to get a larger objective value while keeping the function feasible for the problem (61). Suppose y_{k-1} is changed into y. To make $\int_{\underline{x}}^{\bar{x}} h(x)dx$ fixed, we must change y_k into

$$\frac{y_{k-1}(a_{k-1} - a_{k-2}) + y_k(a_k - a_{k-1}) - (a_{k-1} - a_{k-2})y}{(a_k - a_{k-1})}.$$

To make the function still non-increasing, y should satisfy

$$y_{k+1} \le \frac{y_{k-1}(a_{k-1} - a_{k-2}) + y_k(a_k - a_{k-1}) - (a_{k-1} - a_{k-2})y}{(a_k - a_{k-1})} \le y \le y_{k-2},$$

i.e.,

$$y \in \left[y_{k-1}\frac{a_{k-1} - a_{k-2}}{a_k - a_{k-2}} + y_k\frac{a_k - a_{k-1}}{a_k - a_{k-2}}, \min\left(y_{k-2}, y_{k-1} + (y_k - y_{k+1})\frac{a_k - a_{k-1}}{a_{k-1} - a_{k-2}} \right) \right].$$

Now the objective value in $(a_{k-2}, a_k]$ becomes

$$H(y) = \int_{a_{k-2}}^{a_{k-1}} (y - g(x))_+ \, dx + \int_{a_{k-1}}^{a_k} \left(\frac{y_{k-1}(a_{k-1} - a_{k-2}) + y_k(a_k - a_{k-1}) - (a_{k-1} - a_{k-2})\, y}{(a_k - a_{k-1})} - g(x) \right)_+ dx.$$

Therefore, by Lemma 12, we have that

$$\begin{aligned} &\int_{a_{k-2}}^{a_{k-1}} (y_{k-1} - g(x))_+ \, dx + \int_{a_{k-1}}^{a_k} (y_k - g(x))_+ \, dx \\ =& H(y_{k-1}) \\ \le& \max\left(H\left(y_{k-1} \frac{a_{k-1} - a_{k-2}}{a_k - a_{k-2}} + y_k \frac{a_k - a_{k-1}}{a_k - a_{k-2}} \right), \right. \\ & \left. H\left(\min\left(y_{k-2}, y_{k-1} + (y_k - y_{k+1}) \frac{a_k - a_{k-1}}{a_{k-1} - a_{k-2}} \right) \right) \right). \end{aligned}$$

By considering which one attains the maximal value, there are three cases.

Case 1:

$$H\left(y_{k-1} \frac{a_{k-1} - a_{k-2}}{a_k - a_{k-2}} + y_k \frac{a_k - a_{k-1}}{a_k - a_{k-2}} \right) > H\left(\min\left(y_{k-2}, y_{k-1} + (y_k - y_{k+1}) \frac{a_k - a_{k-1}}{a_{k-1} - a_{k-2}} \right) \right).$$

In this case,

$$H(y_{k-1}) \le H\left(y_{k-1} \frac{a_{k-1} - a_{k-2}}{a_k - a_{k-2}} + y_k \frac{a_k - a_{k-1}}{a_k - a_{k-2}} \right),$$

which means we can get a better objective value by changing both y_{k-1} and y_k into

$$y_{k-1} \frac{a_{k-1} - a_{k-2}}{a_k - a_{k-2}} + y_k \frac{a_k - a_{k-1}}{a_k - a_{k-2}}.$$

In other words, the new function

$$h(x) = \begin{cases} y_1, & a_0 < x \le a_1 \\ \dots & \\ y_{k-2}, & a_{k-3} < x \le a_{k-2} \\ y_{k-1} \frac{a_{k-1} - a_{k-2}}{a_k - a_{k-2}} + y_k \frac{a_k - a_{k-1}}{a_k - a_{k-2}}, & a_{k-2} < x \le a_k \\ y_{k+1}, & a_k < x \le a_{k+1} \end{cases}$$

has a better objective value than the original function $h_1(x)$. Since the number of steps of $h(x)$ is k, by the induction hypothesis, there is a feasible solution h_2 to the problem (61) s.t.

$$\int_{\underline{x}}^{\bar{x}} (h(x) - g(x))_+ dx \le \int_{\underline{x}}^{\bar{x}} (h_2(x) - g(x))_+ dx,$$

which implies

$$\int_{\underline{x}}^{\bar{x}} (h_1(x) - g(x))_+ dx \le \int_{\underline{x}}^{\bar{x}} (h_2(x) - g(x))_+ dx.$$

So (62) holds in this case.

Case 2:

$$H\left(y_{k-1}\frac{a_{k-1}-a_{k-2}}{a_k-a_{k-2}} + y_k\frac{a_k-a_{k-1}}{a_k-a_{k-2}}\right) \le H\left(\min\left(y_{k-2}, y_{k-1} + (y_k - y_{k+1})\frac{a_k-a_{k-1}}{a_{k-1}-a_{k-2}}\right)\right),$$

and

$$\min\left(y_{k-2}, y_{k-1} + (y_k - y_{k+1})\frac{a_k-a_{k-1}}{a_{k-1}-a_{k-2}}\right) = y_{k-2}.$$

This is to say

$$H(y_{k-1}) \le H(y_{k-2}),$$

and

$$y_k + (y_{k-1} - y_{k-2})\frac{a_{k-1}-a_{k-2}}{a_k-a_{k-1}} \ge y_{k+1},$$

which means we can get a better objective value by changing y_{k-1} into y_{k-2}. In other words, the new function

$$h(x) = \begin{cases} y_1, & a_0 < x \le a_1 \\ \cdots & \\ y_{k-2}, & a_{k-3} < x \le a_{k-1} \\ y_k + (y_{k-1} - y_{k-2})\frac{a_{k-1}-a_{k-2}}{a_k-a_{k-1}}, & a_{k-1} < x \le a_k \\ y_{k+1}, & a_k < x \le a_{k+1} \end{cases}$$

has a better objective value than the original function $h_1(x)$. We can see the number of steps of $h(x)$ is k. By the same argument in case 1, (62) also holds in this case.

Case 3:

$$H\left(y_{k-1}\frac{a_{k-1}-a_{k-2}}{a_k-a_{k-2}} + y_k\frac{a_k-a_{k-1}}{a_k-a_{k-2}}\right) \le H\left(\min\left(y_{k-2}, y_{k-1} + (y_k - y_{k+1})\frac{a_k-a_{k-1}}{a_{k-1}-a_{k-2}}\right)\right),$$

and

$$\min\left(y_{k-2}, y_{k-1} + (y_k - y_{k+1})\frac{a_k-a_{k-1}}{a_{k-1}-a_{k-2}}\right) = y_{k-1} + (y_k - y_{k+1})\frac{a_k-a_{k-1}}{a_{k-1}-a_{k-2}}.$$

This is to say

$$H(y_{k-1}) \le H\left(y_{k-1} + (y_k - y_{k+1})\frac{a_k - a_{k-1}}{a_{k-1} - a_{k-2}}\right),$$

and

$$y_{k-2} \ge y_{k-1} + (y_k - y_{k+1})\frac{a_k - a_{k-1}}{a_{k-1} - a_{k-2}},$$

which means we can get a better objective value by changing y_{k-1} into $y_{k-1} + (y_k - y_{k+1})(a_k - a_{k-1})/(a_{k-1} - a_{k-2})$ and changing y_k into y_{k+1}. In other words, the new function

$$h(x) = \begin{cases} y_1, & a_0 < x \le a_1 \\ \dots & \\ y_{k-2}, & a_{k-3} < x \le a_{k-2} \\ y_{k-1} + (y_k - y_{k+1})\frac{a_k - a_{k-1}}{a_{k-1} - a_{k-2}}, & a_{k-2} < x \le a_{k-1} \\ y_{k+1}, & a_{k-1} < x \le a_{k+1} \end{cases}$$

has a better objective value than the original function $h_1(x)$. We can see the number of steps of $h(x)$ is k. By the same argument in case 1, (62) also holds in this case.

Combining all the cases, we can see (62) holds when $n = k + 1$. By induction, we can see (62) holds for any feasible solution h_1 to the problem (59), which implies the optimal value of the problem (61) is the same as that of the problem (59).

The last thing we need to show is the optimal solution to the problem (61) exists. Suppose we have a sequence of feasible solutions $\{h_n, n \ge 1\}$ to the problem (61) whose objective values converge to the optimal value. Since each h_n is characterized by five bounded real numbers, we can find a subsequence along which the five sequences of real numbers converge. This means we can find an a.e. convergent subsequence of $\{h_n, n \ge 1\}$. By the dominated converge theorem, we can see the limit function is also feasible to the problem (61) (we may need to change its values at the endpoints of the steps to make it left-continuous) and its objective value is exactly the optimal value. Thus, the limit function is the optimal solution to the problem (61). □

Proof of Corollary 3. By Lemma 8, we know that each h_i^* is a step function with at most three steps. In total we have $n_{R_0} + 1 \le 4n + 1$ subproblems and h^* is just a "combination" of these h_i^*'s. Therefore, h^* is a non-increasing left-continuous step function with at most $3(4n+1)$ steps and can be represented by (34) for some (z, w), where z is a $(12n+3)$-dimensional vector that represents "step size" and w is also a $(12n+3)$-dimensional vector that represents "height of each step". Consequently, $\tilde{R}$ defined by

$$\tilde{R} = \{(x, y) \in \mathcal{D}_0 : y_0 \le y \le h^*(x; z, w)\} \tag{63}$$

is a dominating OU set of R_0 by Lemma 7. Here the requirement $(z,w) \in (0,\infty)^{12n+4} \times [0,\infty)^{12n+2}$ is due to $\lambda(\tilde{R}) = \lambda(R_0) > 0$. Note that the number of steps $12n+3$ doesn't rely on the choice of R_0 so the dominating OU set of any closed OU set R_0 with $\lambda(R_0) \in (0,\infty), R_0^X \in (0,\infty), R_0^Y \in (0,\infty)$ can be represented by (63) with possibly different (z,w). In other words, all dominating OU sets are contained in the class $\mathcal{R}^* = \{R_{z,w} : R_{z,w} = \{(x,y) \in \mathcal{D}_0 : y_0 \le y \le h^*(x;z,w)\}$ with h^* defined in (34)$\}$. □

Proof of Proposition 5. By the construction of dominating OU sets, in problem (26), we can restrict the choice of R_s to $\mathcal{R}^*$ which is fully parameterized by (z,w). Further, the cardinality of $(z,w) \in (0,\infty)^{12n+4} \times [0,\infty)^{12n+2}$ and the cardinality of $s \in (0,\infty)$ are both continuum. So the distribution $G(s)$ on the index s in problem (26) is equivalent to a distribution Q on (z,w). Thus, problem (26) can be rewritten as the following problem:

$$\begin{aligned}
\max\ & c \int \frac{\lambda(S \cap R_{z,w})}{\lambda(R_{z,w})} dQ(z,w) \\
\text{subject to}\ & \int \frac{R^Y_{z,w}}{\lambda(R_{z,w})} dQ(z,w) \le \frac{u_X}{c} \\
& \int \frac{R^X_{z,w}}{\lambda(R_{z,w})} dQ(z,w) \le \frac{u_Y}{c} \\
& a_i \le \int \frac{\lambda(\{(x,y) : x_{1i} \le x \le x_{2i}, y_{1i} \le y \le y_{2i}\} \cap R_{z,w})}{\lambda(R_{z,w})} dQ(z,w) \le b_i, i = 1,\dots,n \\
& \text{all the integrands are measurable,}
\end{aligned} \tag{64}$$

where $R_{z,w}$ is the closed OU set represented by $R_{z,w} = \{(x,y) \in \mathcal{D}_0 : y_0 \le y \le h^*(x;z,w)\}$ with h^* defined in (34). In view of the form of h^* and the representation of $S = \{(x,y) \in \mathcal{D}_0 : y \ge g(x)\}$ for some known function $g : [x_0,\infty) \mapsto [y_0,\infty]$, we can obtain

$$\lambda(R_{z,w}) = \sum_{i=1}^{12n+3} \sum_{j=1}^{12n+4-i} z_i w_j, \quad R^Y_{z,w} = \sum_{i=1}^{12n+3} w_i, \quad R^X_{z,w} = \sum_{i=1}^{12n+3} z_i,$$

$$\lambda(S \cap R_{z,w}) = \sum_{i=1}^{12n+3} \int \left(y_0 + \sum_{j=1}^{12n+4-i} w_j - g(x) \right)_+ I\left(x_0 + \sum_{j=1}^{i-1} z_j < x \le x_0 + \sum_{j=1}^{i} z_j \right) dx,$$

and

$$\begin{aligned}
&\lambda(\{(x,y): x_{1i} \le x \le x_{2i}, y_{1i} \le y \le y_{2i}\} \cap R_{z,w}) \\
&= \sum_{i=1}^{12n+3} \int I\left(\sum_{j=1}^{i-1} z_j < x - x_0 \le \sum_{j=1}^{i} z_j, x_{1k} \le x \le x_{2k}\right) \\
&\times \left(\min\left(y_0 + \sum_{j=1}^{12n+4-i} w_j, y_{2k}\right) - \min\left(y_0 + \sum_{j=1}^{12n+4-i} w_j, y_{1k}\right)\right) dx.
\end{aligned}$$

Plugging them into problem (64) and noticing that "all the integrands are measurable" automatically holds, we get the moment problem (35). □

Proof of Theorem 2. (1): By Corollary 2 and Proposition 5, we know the optimal value of the OU-DRO problem (25) is not greater than the optimal value of the moment problem (35). On the other hand, any feasible solution to the moment problem (35) can be transformed into a feasible solution to the OU-DRO problem (25) with the same objective value by means of the OU distribution Q defined in (36). Combining two directions together, we know that optimal values of the OU-DRO problem (25) and the moment problem (35) are the same. The equivalence of the moment problem (35) and the non-linear optimization (37) is ensured by Theorem 3.2 in [95] and the conditions in Theorem 3.2 can be verified by Theorem 2.1 and Proposition 3.1 in the same paper. The optimality of P^* is clear since it has the same objective value as the optimal solution Q^*.

(2): By part (1), the OU-DRO problem (25) has the same optimal value as the non-linear optimization (37). Thus, when the constraint $\bar{F}(x_0, y_0) = c$ is replaced by $l_{\bar{F}} \le \bar{F}(x_0, y_0) \le u_{\bar{F}}$, it suffices to make c vary in the interval $[l_{\bar{F}}, u_{\bar{F}}]$ when solving the moment problem (35) or the non-linear optimization (37). Thus, these problems have the same optimal value. Finally, P^* is optimal since its objective value is exactly the optimal value of both the modified moment problem and the DRO problem. □

Proof of Lemma 9. The proof of $\bar{K}$ and $\bar{K}^\circ$ being OU is similar to the proof of Lemma 2 and thus is omitted. To show K is Lebesgue measurable and satisfies $\lambda(K^\circ) = \lambda(K) = \lambda(\bar{K})$, it suffices to show $\lambda(\partial K) = 0$. Without loss of generality, assume x_0 is the origin. Let $O_1, \ldots, O_{2^d}$ be the 2^d (closed) orthants of $\mathbb{R}^d$. We have

$$K = \bigcup_{i=1}^{2^d} (K \cap O_i)$$

and thus

$$\partial K \subset \bigcup_{i=1}^{2^d} \partial(K \cap O_i).$$

Note that by suitable reflection, each $K \cap O_i$ can be transformed into an OU set about the origin on the first orthant $[0,\infty)^d$. Then by [66], we have $\lambda(\partial(K \cap O_i)) = 0$ for each $i = 1, \ldots, 2^d$, which implies $\lambda(\partial K) = 0$. □

Proof of Theorem 3. The first "if and only if" claim can be proved by the similar arguments in the proof of Theorem 1. The second "if and only if" claim is obvious. To prove the Choquet representation, we first note that

$$P(B) = \sum_{1 \le i \le 2^d, P(O_i) > 0} P(O_i) P(B|O_i)$$

by the law of total probability. By Theorem 1, $P(B|O_i)$ can be written as

$$P(B|O_i) = \int_0^\infty W_{\bar{C}_s^i}(B) g_i(s) ds,$$

which justifies (42). Note that $\bar{C}_s^i$ is an OU set which is fully contained in O_i. To show (42) is indeed a Choquet representation, it suffices to show $W_{\bar{C}_s^i}$ is an extreme point in the class of OU distributions about the origin if $\lambda(\bar{C}_s^i) > 0$. This is proved in Lemma 10. □

Proof of Lemma 10. The proof is similar to the proof of Lemma 3 and thus is omitted. □

Proof of Theorem 4. Suppose there is a density $f(x)$ of P such that for every $s > 0$, the set

$$C_s = \{x \in \mathcal{D}_0 : f(x) \ge s\}$$

is d'-POU about x_0. Write W_{C_s} as the uniform distribution on C_s and let $g(s) = \lambda(C_s) \ge 0$. Notice that

$$f(x) = \int_0^{f(x)} 1 ds = \int_0^\infty I(x \in C_s) ds.$$

Thus for any measurable set B,

$$P(B) = \int_B f(x) dx = \int_B \int_0^\infty I(x \in C_s) ds dx = \int_0^\infty \lambda(B \cap C_s) ds = \int_0^\infty W_{C_s}(B) g(s) ds.$$

Setting $B = \mathcal{D}_0$, we can see $P(B) = 1 = \int_0^\infty g(s) ds$ and thus $g(s)$ is a probability density on $(0,\infty)$. Since the probability measure with the density $g(s)$ must be the limit of a sequence of discrete probability measures with finite support, we can see P is in the closed convex hull of the set of all uniform distributions on d'-POU subsets of $\mathcal{D}_0$, i.e., P is a d'-POU distribution.

To prove the "only if" part, we assume P is d'-POU about x_0. Suppose that Q is the uniform distribution on a d'-POU set $K \subset \mathcal{D}_0$ and $0 < \lambda(K) < \infty$. For a point $x \in \mathbb{R}^d$ and $\delta > 0$, we define the neighborhood $N_\delta(x) = \{y \in \mathbb{R}^d : \max_{1\le i\le d} |x_i - y_i| < \delta\}$ as in the proof of Theorem 1. For $x, x' \in \mathcal{D}_0^\circ$ with $x_i \ge x_i'$ for $i = 1, \dots, d'$, $x_i = x_i'$ for $i = d'+1, \dots, d$ and $0 < \delta < \min_{1\le i\le d'}(x_i' - x_{i0})$, we have

$$y \in K \cap N_\delta(x) \Rightarrow y + x' - x \in K \cap N_\delta(x').$$

Since Lebesgue measure is translation invariant, we can get

$$\lambda(K \cap N_\delta(x')) \ge \lambda(K \cap N_\delta(x)).$$

Dividing by $\lambda(K)$, we have

$$Q(N_\delta(x')) \ge Q(N_\delta(x)). \tag{65}$$

Clearly, this relation also holds under the convex combinations of the uniform distributions on d'-POU sets. Since P is d'-POU about x_0, by definition, there is a sequence $\{Q_m, m \ge 1\}$ such that Q_m converges weakly to P, where Q_m's are the convex combinations of the uniform distributions on d'-POU sets. Therefore, these Q_m's satisfy (65). Since P has a density, say f_0, we have

$$P(\partial N_\delta(x')) = P(\partial N_\delta(x)) = 0.$$

Weak convergence and (65) imply that

$$P(N_\delta(x')) \ge P(N_\delta(x)). \tag{66}$$

For $x \in \mathcal{D}_0^\circ$, we define

$$f(x) = \limsup_{\delta \downarrow 0} \frac{P(N_\delta(x))}{\lambda(N_\delta(x))} = \limsup_{\delta \downarrow 0} \frac{P(N_\delta(x))}{(2\delta)^d}.$$

By the Lebesgue differentiation theorem, $f(x) = f_0(x)$ a.e., which means $f(x)$ is also a density for P. Besides, (66) implies $f(x') \ge f(x)$ for $x, x' \in \mathcal{D}_0^\circ$ with $x_i \ge x_i'$ for $i = 1, \dots, d'$, $x_i = x_i'$ for $i = d'+1, \dots, d$. For $x \in \partial\mathcal{D}_0$, we simply define

$$f(x) = \sup_{y \in \mathcal{D}_0^\circ} f(y).$$

Then the density $f(x)$ is d'-POU about x_0 on $\mathcal{D}_0$.

The equivalence of C_s being d'-POU about x_0 and f being a d'-POU density about x_0 on $\mathcal{D}_0$ is easy to see by definition. □

Proof of Theorem 5. From the proof of Theorem 4, we have

$$P(B) = \int_0^\infty \lambda(C_s \cap B) ds. \tag{67}$$

By Fubini's theorem, we have

$$\begin{aligned}
\lambda(C_s \cap B) &= \int_{\mathcal{D}_0} I(x \in C_s \cap B) dx_1 \cdots dx_d \\
&= \int_{x_{d'+1,0}}^\infty \cdots \int_{x_{d0}}^\infty \left(\int_{x_{10}}^\infty \cdots \int_{x_{d'0}}^\infty I(x \in C_s \cap B) dx_1 \cdots dx_{d'} \right) dx_{d'+1} \cdots dx_d \\
&= \int_{x_{d'+1,0}}^\infty \cdots \int_{x_{d0}}^\infty \lambda_{d'}((C_s \cap B)_{x_{d'+1},\ldots,x_d}) dx_{d'+1} \cdots dx_d \\
&= \int_{x_{d'+1,0}}^\infty \cdots \int_{x_{d0}}^\infty \lambda_{d'}(C_{s,x_{d'+1},\ldots,x_d} \cap B_{x_{d'+1},\ldots,x_d}) dx_{d'+1} \cdots dx_d \\
&= \int_{x_{d'+1,0}}^\infty \cdots \int_{x_{d0}}^\infty W_{C_{s,x_{d'+1},\ldots,x_d}}(B_{x_{d'+1},\ldots,x_d}) \lambda_{d'}(C_{s,x_{d'+1},\ldots,x_d}) dx_{d'+1} \cdots dx_d.
\end{aligned}$$

Notice that for any set K we have

$$K_{x_{d'+1},\ldots,x_d} \times \{(x_{d'+1},\ldots,x_d)\} = K \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1},\ldots,y_d = x_d\}.$$

Therefore

$$\begin{aligned}
&W_{C_{s,x_{d'+1},\ldots,x_d}}(B_{x_{d'+1},\ldots,x_d}) \\
=&W_{C_s \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1},\ldots,y_d = x_d\}}(B \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1},\ldots,y_d = x_d\}) \\
=&W_{C_s \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1},\ldots,y_d = x_d\}}(B).
\end{aligned}$$

It follows that $\lambda(C_s \cap B)$ can be represented as

$$\lambda(C_s \cap B) = \int_{x_{d'+1,0}}^\infty \cdots \int_{x_{d0}}^\infty W_{C_s \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1},\ldots,y_d = x_d\}}(B) \lambda_{d'}(C_{s,x_{d'+1},\ldots,x_d}) dx_{d'+1} \cdots dx_d. \tag{68}$$

Plugging (68) into (67), we get

$$\begin{aligned}
&P(B) \\
=&\int_0^\infty \int_{x_{d'+1,0}}^\infty \cdots \int_{x_{d0}}^\infty W_{C_s \cap \{y \in \mathbb{R}^d : y_{d'+1} = x_{d'+1},\ldots,y_d = x_d\}}(B) g(s, x_{d'+1},\ldots,x_d) dx_{d'+1} \cdots dx_d ds,
\end{aligned} \tag{69}$$

where $g(s,x_{d'+1},\dots,x_d)=\lambda_{d'}(C_{s,x_{d'+1},\dots,x_d})$. Letting $B=\mathcal{D}_0$, we can see $g(s,x_{d'+1},\dots,x_d)$ is a probability density. Since only $W_{C_s\cap\{y\in\mathbb{R}^d:y_{d'+1}=x_{d'+1},\dots,y_d=x_d\}}$ with $\lambda_{d'}(C_{s,x_{d'+1},\dots,x_d})>0$ contributes to (69), in order to prove (69) is indeed a Choquet representation, it suffices to show $W_{C_s\cap\{y\in\mathbb{R}^d:y_{d'+1}=x_{d'+1},\dots,y_d=x_d\}}$ with $\lambda_{d'}(C_{s,x_{d'+1},\dots,x_d})>0$ is an extreme point in the class of d'-POU distributions. This is proved in Lemma 11. □

Proof of Lemma 11. Suppose there exist two d'-POU distributions P_1, P_2 and $\eta\in(0,1)$ s.t.

$$W_K=\eta P_1+(1-\eta)P_2.$$

Since the support of W_K is contained in $\{y\in\mathbb{R}^d:y_{d'+1}=x_{d'+1},\dots,y_d=x_d\}$, the supports of P_1 and P_2 are also contained in $\{y\in\mathbb{R}^d:y_{d'+1}=x_{d'+1},\dots,y_d=x_d\}$. Then we can reduce W_K,P_1 and P_2 to OU distributions on the subspace $\{y\in\mathbb{R}^d:y_{d'+1}=x_{d'+1},\dots,y_d=x_d\}$. Note that W_K is reduce to $W_{K_{x_{d'+1},\dots,x_d}}$ on $\{x\in\mathbb{R}^{d'}:x_i\geq x_{i0},i=1,\dots,d'\}$ and $K_{x_{d'+1},\dots,x_d}$ is an OU set about $(x_{10},\dots,x_{d'0})$ on $\{x\in\mathbb{R}^{d'}:x_i\geq x_{i0},i=1,\dots,d'\}$ with $\lambda_{d'}(K_{x_{d'+1},\dots,x_d})>0$. By Lemma 3, $W_{K_{x_{d'+1},\dots,x_d}}$ is an extreme point in the class of OU distributions about $(x_{10},\dots,x_{d'0})$ on $\{x\in\mathbb{R}^{d'}:x_i\geq x_{i0},i=1,\dots,d'\}$. Therefore, the distributions of P_1 and P_2 on $\{y\in\mathbb{R}^d:y_{d'+1}=x_{d'+1},\dots,y_d=x_d\}$ are the same as $W_{K_{x_{d'+1},\dots,x_d}}$, which means P_1 and P_2 are equal to W_k. Therefore, W_K is an extreme point in the class of d'-POU distributions about x_0 on $\mathcal{D}_0$. □

Proof of Theorem 6. We use the similar arguments for star unimodality in [30]. We first prove the "if" part. Suppose $X\overset{d}{=}(x_{10}+U(Z_1-x_{10}),Z_2,\dots,Z_d)$. We need to prove X is 1-POU. In fact, we only need to prove it when $(Z_1,\dots,Z_d)$ is degenerate, i.e., $(Z_1,\dots,Z_d)$ put mass 1 at a point $(z_1,\dots,z_d)\in\mathcal{D}_0$. Then by taking the convex mixture and weak limit and noticing that 1-POU distribution is closed under these operations, we can see it holds for any $(Z_1,\dots,Z_d)$. When $(Z_1,\dots,Z_d)$ is degenerate at $(z_1,\dots,z_d)\in\mathcal{D}_0$, $X\overset{d}{=}(x_{10}+U(z_1-x_{10}),z_2,\dots,z_d)$ whose distribution is uniform on $[x_{10},z_1]\times\{(z_2,\dots,z_d)\}$. Since $[x_{10},z_1]\times\{(z_2,\dots,z_d)\}$ is a 1-POU set on $\mathcal{D}_0$, X is clearly 1-POU. This proves the "if" part.

Now we consider the "only if" part. Suppose $X\equiv(X_1,\dots,X_d)$ is the uniform distribution on a 1-POU set K which can be written as

$$K=\{x\equiv(x_1,\dots,x_d)\in\mathcal{D}_0:x_{10}\leq x_1\leq g(x_2,\dots,x_d)\},$$

where $g(x_2,\dots,x_d)>x_{10}$ is continuous with domain $Dom(g)=[x_{20},\infty)\times\cdots\times[x_{d0},\infty)$ and $\lambda(K)\in(0,\infty)$. Then the density of X is

$$f_X(x)=\frac{1}{\int_{Dom(g)}(g(x_2,\dots,x_d)-x_{10})dx_2\cdots dx_d}I(x_{10}\leq x_1\leq g(x_2,\dots,x_d)).$$

Therefore, the conditional density of X_1 given $(X_2,\dots,X_d)=(x_2,\dots,x_d)$ is

$$f_{X_1|(X_2,\dots,X_d)}(x_1|x_2,\dots,x_d)=\frac{1}{g(x_2,\dots,x_d)-x_{10}}I(x_{10}\le x_1\le g(x_2,\dots,x_d)).$$

Let $U=(X_1-x_{10})/(g(X_2,\dots,X_d)-x_{10})$. Then we can see the conditional distribution of U given $(X_2,\dots,X_d)=(x_2,\dots,x_d)$ is the uniform distribution on $(0,1)$ and thus is independent of $(X_2,\dots,X_d)$. So X can be represented by

$$\begin{aligned}X&\equiv(X_1,\dots,X_d)\\&=(x_{10}+U(g(X_2,\dots,X_d)-x_{10}),X_2,\dots,X_d)\\&=(x_{10}+U(Z_1-x_{10}),Z_2,\dots,Z_d),\end{aligned}$$

where $(Z_1,Z_2,\dots,Z_d)=(g(X_2,\dots,X_d),X_2,\dots,X_d)$ and U is independent of $(Z_1,Z_2,\dots,Z_d)$. By taking the convex mixture and weak limit, we can see "only if" part holds for any X.

The representation (44) is just a reformulation of $X=(x_{10}+U(Z_1-x_{10}),Z_2,\dots,Z_d)$ by conditioning on the values of $(Z_1,\dots,Z_d)$. By Lemma 11, we know $W_{1\text{-POU}}(z)$ is an extreme point in the class of 1-POU distributions about x_0 on $\mathcal{D}_0$. So (44) is indeed a Choquet representation. Next we will show Q is uniquely determined by P, i.e., the Choquet representation (44) is unique. We write φ_X and φ_Z as the characteristic function of X and Z respectively. For $t\in\mathbb{R}^n$, we have

$$\begin{aligned}\varphi_X(t)&=E[e^{it^\top X}]=E[e^{it^\top(x_{10}+U(Z_1-x_{10}),Z_2,\dots,Z_d)}]=\int_0^1 E[e^{it^\top(x_{10}+u(Z_1-x_{10}),Z_2,\dots,Z_d)}]du\\&=\int_0^1 e^{it_1x_{10}(1-u)}\varphi_Z(ut_1,t_2,\dots,t_d)du=e^{it_1x_{10}}\int_0^1 e^{-it_1ux_{10}}\varphi_Z(ut_1,t_2,\dots,t_d)du.\end{aligned}$$

So for $y>0$, we have

$$\varphi_X(yt_1,t_2,\dots,t_d)=e^{iyt_1x_{10}}\int_0^1 e^{-it_1yux_{10}}\varphi_Z(yut_1,t_2,\dots,t_d)du$$

By change of variable $v=uy$, we have

$$\begin{aligned}&\varphi_X(yt_1,t_2,\dots,t_d)=\frac{e^{iyt_1x_{10}}}{y}\int_0^y e^{-it_1vx_{10}}\varphi_Z(vt_1,t_2,\dots,t_d)dv\\\Leftrightarrow&\frac{y}{e^{iyt_1x_{10}}}\varphi_X(yt_1,t_2,\dots,t_d)=\int_0^y e^{-it_1vx_{10}}\varphi_Z(vt_1,t_2,\dots,t_d)dv.\end{aligned}\tag{70}$$

Since the right hand side of (70) is differentiable with respect to y, so is the left hand side. Then we have

$$\begin{aligned}&\frac{d}{dy}\left[\frac{y}{e^{iyt_1x_{10}}}\varphi_X(yt_1,t_2,\dots,t_d)\right]=e^{-it_1yx_{10}}\varphi_Z(yt_1,t_2,\dots,t_d)\\ \Leftrightarrow &e^{it_1yx_{10}}\frac{d}{dy}\left[\frac{y}{e^{iyt_1x_{10}}}\varphi_X(yt_1,t_2,\dots,t_d)\right]=\varphi_Z(yt_1,t_2,\dots,t_d).\end{aligned}$$

Letting $y=1$, we can see φ_X determines φ_Z and hence P determines Q.

Finally, suppose $P(X_1=x_{10})=0$, we need to prove X_1 is an absolutely continuous random variable. Note that $X_1\overset{d}{=}x_{10}+U(Z_1-x_{10})$ so we have $P(Z_1=x_{10})=0$. Consider any measurable set A with $\lambda(A)=0$ and write Q_{Z_1} as the probability distribution of Z_1. We have

$$\begin{aligned}&P(X_1\in A)\\ =&P(x_{10}+U(Z_1-x_{10})\in A)\\ =&\int_{(x_{10},\infty)}P(x_{10}+U(z_1-x_{10})\in A)dQ_{Z_1}(z_1)\\ =&\int_{(x_{10},\infty)}0dQ_{Z_1}(z_1)\\ =&0,\end{aligned}$$

which means X_1 is absolutely continuous. Now we prove $E_Q[I(Z_1 \geq x)/(Z_1 - x_{10})], x \geq x_{10}$ is a probability density of X_1. For any $x_1 > x_{10}$, we have

$$\begin{aligned}
&P(X_1 \geq x_1)\\
=&P(x_{10} + U(Z_1 - x_{10}) \geq x_1)\\
=&\int_{[x_1,\infty)} P(x_{10} + U(z_1 - x_{10}) \geq x_1) dQ_{Z_1}(z_1)\\
=&\int_{[x_1,\infty)} \frac{z_1 - x_1}{z_1 - x_{10}} dQ_{Z_1}(z_1)\\
=&\int_{[x_1,\infty)} \int_{[x_1,z_1]} \frac{1}{z_1 - x_{10}} dx dQ_{Z_1}(z_1)\\
=&\int_{[x_1,\infty)} \int_{[x,\infty)} \frac{1}{z_1 - x_{10}} dQ_{Z_1}(z_1) dx\\
=&\int_{[x_1,\infty)} E_Q\left[\frac{I(Z_1 \geq x)}{Z_1 - x_{10}}\right] dx.
\end{aligned}$$

Since this holds for any $x_1 > x_{10}$, we know $E_Q[I(Z_1 \geq x)/(Z_1 - x_{10})], x \geq x_{10}$ is indeed a probability density of X_1. Moreover, by Monotone Convergence Theorem and $P(Z_1 = x_{10}) = 0$,

$$\lim_{x \downarrow x_{10}} E_Q\left[\frac{I(Z_1 \geq x)}{Z_1 - x_{10}}\right] = E_Q\left[\lim_{x \downarrow x_{10}} \frac{I(Z_1 \geq x)}{Z_1 - x_{10}}\right] = E_Q\left[\frac{I(Z_1 > x_{10})}{Z_1 - x_{10}}\right] = E_Q\left[\frac{I(Z_1 \geq x_{10})}{Z_1 - x_{10}}\right],$$

which means $E_Q[I(Z_1 \geq x)/(Z_1 - x_{10})]$ is continuous at $x = x_{10}$. □

Proof of Theorem 7. Without loss of generality suppose P is the probability measure induced by X. Notice that we have $\bar{F}(x_0) = c$ in (46). Therefore, P/c is a 1-POU probability measure on $\mathcal{D}_0$. By Theorem 6, we have the representation

$$P \equiv c\frac{P}{c} = c\int_{\mathcal{D}_0} W_{\text{1-POU}}(z) dQ(z) \tag{71}$$

for some probability measure Q on $\mathcal{D}_0$. We write $Z = (Z_1, \ldots, Z_d)$ as a random vector with the distribution Q and write U as the uniform distribution on $(0, 1)$ independent of Z. Then $(x_{10} + U(Z_1 - x_{10}), Z_2, \ldots, Z_d)$ has the probability distribution P/c on $\mathcal{D}_0$.

We first consider the reduction of the objective function. Since S is in the form of (47), by (71), we have

$$\begin{aligned}
&P((X_1,\dots,X_d)\in S)\\
\equiv&P(S)\\
=&c\int_{\mathcal{D}_0}\frac{\min(g_2(z_2,\dots,z_d),z_1)-\min(g_1(z_2,\dots,z_d),z_1)}{z_1-x_{10}}dQ(z)\\
=&cE_Q\left[\frac{\min(g_2(Z_2,\dots,Z_d),Z_1)-\min(g_1(Z_2,\dots,Z_d),Z_1)}{Z_1-x_{10}}\right].
\end{aligned}$$

So the objective function is expressed by the expectation under Q.

Next we consider the density constraint $f_{X_1}(x_{10})\le u_{X_1}$. Since f_{X_1} is the marginal density of X_1 within $\mathcal{D}_0$, we can see f_{X_1}/c is the probability density of $x_{10}+U(Z_1-x_{10})$. Recall that we require f_{X_1} to be continuous at $x=x_{10}$. So the value $f_{X_1}(x_{10})$ is uniquely determined and is actually given by Theorem 6, i.e.,

$$f_{X_1}(x_{10})=c\frac{f_{X_1}(x_{10})}{c}=cE_Q\left[\frac{I(Z_1\ge x_{10})}{Z_1-x_{10}}\right]=cE_Q\left[\frac{1}{Z_1-x_{10}}\right],$$

where the last inequality follows from the fact that Z_1 takes values in $[x_{10},\infty)$. Therefore, the density constraint can be rewritten as

$$E_Q\left[\frac{1}{Z_1-x_{10}}\right]\le\frac{u_{X_1}}{c}.$$

Moreover, this constraint also implies that $Q(Z_1=x_{10})=0$ for any feasible probability measure Q. Therefore, $P(X_1=x_{10})=P(x_{10}+U(Z_1-x_{10})=x_{10})=0$. In other words, the constraint that $P(X_1=x_{10})=0$ is already involved in the density constraint.

Then we consider the moment constraint

$$a_i\bar{F}(x_0)\le P(\underline{x}_{ji}\le X_j\le\bar{x}_{ji},j=1,\dots,d)\le b_i\bar{F}(x_0).$$

By (71) and $\bar{F}(x_0)=c$, it can be represented by

$$a_i\le\int_{\mathcal{D}_0}\frac{\min(z_1,\bar{x}_{1i})-\min(z_1,\underline{x}_{1i})}{z_1-x_{10}}I(\underline{x}_{ji}\le z_j\le\bar{x}_{ji},j\ge2)dQ(z)\le b_i,$$

i.e.,

$$a_i\le E_Q\left[\frac{\min(Z_1,\bar{x}_{1i})-\min(Z_1,\underline{x}_{1i})}{Z_1-x_{10}}I(\underline{x}_{ji}\le Z_j\le\bar{x}_{ji},j\ge2)\right]\le b_i.$$

Finally, by Theorem 6, the representation (71) is an "if and only if" condition for P being 1-POU about x_0 on $\mathcal{D}_0$. Hence, the moment problem (48) is equivalent to the 1-POU-DRO problem (46). Moreover, the optimal solution of the two problems are also related by the representation given in Theorem 6. This completes our proof. □

Proof of Corollary 4. This follows directly from Theorem 3.2 in [95]. □