Online Appendix to "Data-Driven Distributionally Robust CVaR Portfolio Optimization under Regime-Switching Ambiguity Set"

EC.1. Proof in Section 3

We first make the following notation. For any given $(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{u}}, \tilde{n}, \tilde{k}) \sim \mathbb{P}^+ \in \mathcal{U}_{rs}(\boldsymbol{\theta})$, we use $\Pi_{\tilde{\boldsymbol{r}}}\mathbb{P}^+$ to represent the distribution of that $\tilde{\boldsymbol{r}}$. And we use $\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta})$ to represent the ambiguity set of $\tilde{\boldsymbol{r}}$ where $(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{u}}, \tilde{n}, \tilde{k}) \in \mathcal{U}_{rs}(\boldsymbol{\theta})$ for some $(\tilde{\boldsymbol{u}}, \tilde{n}, \tilde{k})$.

Proof of Proposition 1. We first consider our ambiguity set under each regime $k \in [K]$ below with each ball center $\hat{\mathbb{P}}_k$.

$$\mathcal{U}_{rs}^{k}(\theta_k) = \left\{ \mathbb{P} \in \mathcal{P}\left(\mathbb{R}^I \times \mathbb{R} \times [N_k]\right) \left| \begin{array}{l} (\tilde{\boldsymbol{r}}, \tilde{u}, \tilde{n}) \sim \mathbb{P} \\ \mathbb{E}^{\mathbb{P}}[\tilde{u}] \le \theta_k \\ \mathbb{P}\left[\tilde{\boldsymbol{r}} \in \mathbb{R}^I, \rho(\tilde{\boldsymbol{r}}, \boldsymbol{r}_{nk}) \le \tilde{u} \mid n \in [N_k]\right] = 1 \\ \mathbb{P}[\tilde{n} = n] = \frac{1}{N_k} \ , \forall n \in [N_k] \end{array} \right. \right\} \tag{EC.1}$$

First we show that $\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta}) \subseteq \mathcal{U}_w(\sum_{k\in[K]} w_k\theta_k)$. That is to say, fix any $\mathbb{P} \in \mathcal{U}_{rs}(\boldsymbol{\theta})$ and we write its projection over $\tilde{\boldsymbol{r}}$ as $\Pi_{\tilde{\boldsymbol{r}}}\mathbb{P} = \sum_{k\in[K]} w_k \mathbb{P}_k = \sum_{k\in[K]} w_k \sum_{n\in[N_k]} \frac{1}{N_k}\mathbb{P}_{nk}$, where $\mathbb{P}_k \in \Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}^k(\theta_k)$ is the conditional distribution of $\tilde{\boldsymbol{r}}$ under regime k and $\mathbb{P}_{nk}$ is the conditional distribution of $\tilde{\boldsymbol{r}}$ given the outcome of random scenario is n under that regime. We can then construct a joint distribution by the natural coupling $\nu = \sum_{k\in[K]} \frac{w_k}{N_k} \sum_{n\in[N_k]} \mathbb{P}_{nk} \otimes \delta_{\boldsymbol{r}_{nk}}$. Recall the definition of Wasserstein distance, we have:

$$W_m(\Pi_{\tilde{\boldsymbol{r}}}\mathbb{P}, \hat{\mathbb{P}}) \le \mathbb{E}^{\nu}[\rho(\tilde{\boldsymbol{r}}, \boldsymbol{r})] = \sum_{k\in[K]} w_k \sum_{n\in[N_k]} \frac{1}{N_k} \mathbb{E}_{\tilde{\boldsymbol{r}}\sim\mathbb{P}_{nk}}[\rho(\tilde{\boldsymbol{r}}, \boldsymbol{r}_{nk})] \le \sum_{k\in[K]} w_k\theta_k.$$

To show the other side, we argue by contradiction. We can construct one $\mathbb{P}$ such that $W_m(\mathbb{P}, \hat{\mathbb{P}}) = \sum_k w_k\theta_k$ but for that $\tilde{\boldsymbol{r}} \sim \mathbb{P}$, there is no associated $(\tilde{\boldsymbol{u}}, \tilde{n}, \tilde{k})$ such that $(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{u}}, \tilde{n}, \tilde{k}) \in \mathcal{U}_{rs}(\boldsymbol{\theta})$. Below is a simple counterexample when $I = 1$.

When $I = 1$, computing Wasserstein distance with any $m \in [1, +\infty]$ would lead to the same result. For simplicity, we take the case when $m = 1, K = 2$ and $w_k = \frac{1}{2}, N_k = 1\ \forall k \in [K]$. We let $\hat{\mathbb{P}} = \frac{1}{2}\delta_{\boldsymbol{r}_{11}} + \frac{1}{2}\delta_{\boldsymbol{r}_{12}}$ and $\mathbb{P}^+ = \frac{1}{2}\delta_{\tilde{\boldsymbol{r}}_{11}} + \frac{1}{2}\delta_{\tilde{\boldsymbol{r}}_{12}}$, where we choose θ_1, θ_2 such that $\tilde{\boldsymbol{r}}_{11} = \boldsymbol{r}_{11} - (\theta_1 + \frac{\theta_2}{2})$ and $\tilde{\boldsymbol{r}}_{12} = \boldsymbol{r}_{12} + \frac{\theta_2}{2}$ with $\tilde{\boldsymbol{r}}_{11} < \boldsymbol{r}_{11} < \boldsymbol{r}_{11} + \theta_1 < \boldsymbol{r}_{12} < \tilde{\boldsymbol{r}}_{12}$. By moving the mass of $\tilde{\boldsymbol{r}}_{11}$ to $\boldsymbol{r}_{11}$ and $\tilde{\boldsymbol{r}}_{12}$ to $\boldsymbol{r}_{12}$, the total transportation costs are $\frac{1}{2}\rho(\tilde{\boldsymbol{r}}_{11}, \boldsymbol{r}_{11}) + \frac{1}{2}\rho(\tilde{\boldsymbol{r}}_{12}, \boldsymbol{r}_{12}) = \frac{\theta_1+\theta_2}{2}$. Therefore, $W_m(\mathbb{P}^+, \hat{\mathbb{P}}) \le \frac{\theta_1+\theta_2}{2}$. However, $\mathbb{P}[\rho(\tilde{\boldsymbol{r}}, \boldsymbol{r}_{11}) \le \theta_1] = 0, \forall \tilde{\boldsymbol{r}} \in \mathbb{P}^+$, which implies $\mathbb{P}^+ \not\subseteq \Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta})$. Therefore, we show that $\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta})$ is a true subset of $\mathcal{U}_w(\sum_{k\in[K]} w_k\theta_k)$. $\square$

As a by-product of this proposition, we also obtain a simple ambiguity set representation of $\Pi_{\tilde{\boldsymbol{r}}}(\boldsymbol{\theta})$ from Wasserstein distance. This result is a regime-based extension of Theorem 2 in Chen et al. (2020). We would use this to prove the measure concentration result later.

LEMMA EC.1. *For all $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$, we have:*

$$\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta}) = \left\{ \mathbb{P} : \mathbb{P} = \sum_{k\in[K]} w_k \mathbb{P}_k, W_m(\mathbb{P}_k, \hat{\mathbb{P}}_k) \le \theta_k \right\}, \tag{EC.2}$$

where $\hat{\mathbb{P}}_k = \frac{1}{N_k}\sum_{n\in[N_k]} \delta_{\boldsymbol{r}_{nk}}$.

Proof of Lemma EC.1. In fact by definition of $\mathcal{U}_{rs}^k(\theta_k)$ and $\mathcal{U}_{rs}(\boldsymbol{\theta})$, we have:

$$\begin{aligned}\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta}) &= \left\{\mathbb{P}|\mathbb{P} := \sum_{k\in[K]} w_k\mathbb{P}_k, \mathbb{P}_k \in \Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}^k(\theta_k), \forall k\in[K]\right\}\\ &= \left\{\mathbb{P}|\mathbb{P} := \sum_{k\in[K]} w_k\mathbb{P}_k, W_m(\mathbb{P}_k,\hat{\mathbb{P}}_k) \le \theta_k, \forall k\in[K]\right\},\end{aligned}$$

where the second equality follows by $\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}^k(\theta_k) = \{\mathbb{P}_k|\boldsymbol{r}\sim\mathbb{P}_k, W_m(\mathbb{P}_k,\hat{\mathbb{P}}_k)\le\theta_k\}$ from Theorem 2 in Chen et al. (2020). □

Proof of Theorem 1. We first consider the same problem:

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})} \mathbb{E}^{\mathbb{P}}[G(\boldsymbol{x},\boldsymbol{r})] = \sup_{(\boldsymbol{r},u_k)\sim\tilde{\mathbb{P}}_{nk}} \quad & \textstyle\sum_{k\in[K]}\sum_{n\in[N_k]} \frac{w_k}{N_k}\mathbb{E}^{\tilde{\mathbb{P}}_{nk}}[G(\boldsymbol{x},\boldsymbol{r})]\\ \text{s.t.} \quad & \textstyle\sum_{n\in[N_k]} \frac{w_k}{N_k}\mathbb{E}^{\tilde{\mathbb{P}}_{nk}}[u_k] \le w_k\theta_k, \forall k\in[K],\\ & \tilde{\mathbb{P}}_{nk}[\boldsymbol{r}\in\mathbb{R}^I, \rho(\boldsymbol{r},\boldsymbol{r}_{nk})\le u_k] = 1, \forall n\in[N_k], k\in[K].\end{aligned}$$

Denote the dual variables of each constraint as β_k $\forall k\in[K]$ and $\hat{\alpha}_{nk}$ $\forall n\in[N_k], k\in[K]$, respectively. We derive the dual problem under the proposed regime-switching ambiguity set as follows:

$$\begin{aligned} &\inf_{\hat{\alpha},\boldsymbol{\beta}} \ \textstyle\sum_{k\in[K]}\sum_{n\in[N_k]}\hat{\alpha}_{nk} + \sum_{k\in[K]}\beta_k w_k\theta_k,\\ &\text{s.t. } \hat{\alpha}_{nk} + \frac{w_k}{N_k}u_k\beta_k \ge \frac{w_k}{N_k}G(\boldsymbol{x},\boldsymbol{r}), \forall k\in[K], n\in[N_k], \boldsymbol{r}\in\mathbb{R}^I, \rho(\boldsymbol{r},\boldsymbol{r}_{nk})\le u_k,\\ &\quad\ \beta_k\ge 0, \forall k\in[K].\\ = &\inf_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ \textstyle\sum_{k\in[K]}\sum_{n\in[N_k]}\frac{w_k}{N_k}\alpha_{nk} + \sum_{k\in[K]}\beta_k w_k\theta_k\\ &\text{s.t. } \alpha_{nk} + u_k\beta_k \ge G(\boldsymbol{x},\boldsymbol{r}), \forall k\in[K], n\in[N_k], \boldsymbol{r}\in\mathbb{R}^I, \rho(\boldsymbol{r},\boldsymbol{r}_{nk})\le u_k,\\ &\quad\ \beta_k\ge 0, \forall k\in[K],\end{aligned}$$

where $\alpha_{nk} = \frac{N_k}{w_k}\hat{\alpha}_{nk}, \forall n\in[N_k], k\in[K]$. By the definition of $G(\boldsymbol{x},\boldsymbol{r})$, we have

$$\begin{aligned} &\inf_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ \textstyle\sum_{k\in[K]}\sum_{n\in[N_k]}\frac{w_k}{N_k}\alpha_{nk} + \sum_{k\in[K]}\beta_k w_k\theta_k\\ &\text{s.t. } \alpha_{nk} + \beta_k u_k \ge -\boldsymbol{x}'\boldsymbol{r} - v, \forall k\in[K], n\in[N_k], \boldsymbol{r}\in\mathbb{R}^I, \rho(\boldsymbol{r},\boldsymbol{r}_{nk})\le u_k,\\ &\quad\ \alpha_{nk} + \beta_k u_k \ge 0, \forall k\in[K], n\in[N_k], \boldsymbol{r}\in\mathbb{R}^I, \rho(\boldsymbol{r},\boldsymbol{r}_{nk})\le u_k,\\ &\quad\ \beta_k\ge 0, \forall k\in[K].\end{aligned} \tag{EC.3}$$

There are infinite constraints in this problem. To get a tractable reformulation, we take the first set of infinity constraints as an example to illustrate how to reformulate it into finite constraints. The constraint is equivalent to $Z_1^{nk} := \inf_{\boldsymbol{r}\in\mathbb{R}^I, \rho(\boldsymbol{r},\boldsymbol{r}_{nk})\le u_k} -\boldsymbol{x}'\boldsymbol{r} + \beta_k u_k \ge -\alpha_{nk} - v, \forall n\in[N_k], k\in[K]$, where Z_1^{nk} can be reformulated as a set of linear constraints. We first show a technical lemma in Boyd and Vandenberghe (2004) for conic optimization.

Lemma EC.2 **(Example 5.12 in Boyd and Vandenberghe (2004))**. *We consider the following conic optimization problem,*

$$\min_{\boldsymbol{p}_{nk},u_k} -\boldsymbol{x}'\boldsymbol{p}_{nk} + \beta_k u_k - \boldsymbol{r}_{nk}'\boldsymbol{x}, \ \ s.t.\ \boldsymbol{G}\begin{pmatrix}\boldsymbol{p}_{nk}\\ u_k\end{pmatrix} = \boldsymbol{f}, \ \begin{pmatrix}\boldsymbol{p}_{nk}\\ u_k\end{pmatrix}\in\bar{K}, \tag{EC.4}$$

where $\bar{K}\subseteq\mathbb{R}^{I+1}$ is a proper cone and $\boldsymbol{G}\in\mathbb{R}^{t\times n}, \boldsymbol{f}\in\mathbb{R}^t$ for some t. Its duality problem is as follows:

$$\max_{\boldsymbol{\gamma}\in\mathbb{R}^t} \boldsymbol{f}'\boldsymbol{\gamma} - \boldsymbol{r}_{nk}'\boldsymbol{x}, \ \ s.t.\ \begin{pmatrix}-\boldsymbol{x}\\ \beta_k\end{pmatrix} - \boldsymbol{G}'\boldsymbol{\gamma}\in\bar{K}^*, \tag{EC.5}$$

where $\bar{K}^* = \{\boldsymbol{y} \mid \boldsymbol{y}'\boldsymbol{x} \geq 0, \forall \boldsymbol{x} \in \bar{K}\}$ *denote the dual cone of* $\bar{K}$*. Furthermore, if there exists* $\begin{pmatrix} \boldsymbol{p}_{nk} \\ u_k \end{pmatrix} \succ_{\bar{K}} \boldsymbol{0}$ *with* $\boldsymbol{G}\begin{pmatrix} \boldsymbol{p}_{nk} \\ u_k \end{pmatrix} = \boldsymbol{f}$*, then strong duality holds.*

If $u_k = 0$, Z_1^{nk} reduces to $-\boldsymbol{r}'_{nk}\boldsymbol{b}_{nk} \geq -\alpha_{nk} - v$. Otherwise, we introduce the auxiliary variable $\boldsymbol{p}_{nk} = \boldsymbol{r} - \boldsymbol{r}_{nk}$. And the left-hand side of Z_1^{nk} can be reformulated as:

$$\min_{\boldsymbol{p}_{nk}, u_k} -\boldsymbol{x}'\boldsymbol{p}_{nk} + \beta_k u_k - \boldsymbol{r}'_{nr}\boldsymbol{x}, \text{ s.t. } \|\boldsymbol{p}_{nk}\|_m \leq u_k. \tag{EC.6}$$

When we consider $\bar{K} = \left\{ \begin{pmatrix} \boldsymbol{p}_{nk} \\ u_k \end{pmatrix} \in \mathbb{R}^{I+1} \mid \|\boldsymbol{p}_{nk}\|_m \leq u_k \right\}$. (EC.6) is a special case of the conic optimization problem in Lemma EC.2 as $\begin{pmatrix} \boldsymbol{p}_{nk} \\ u_k \end{pmatrix} \in \bar{K}$. For $\begin{pmatrix} \boldsymbol{p}_{nk} \\ u_k \end{pmatrix} \in \bar{K}$, we have $\begin{pmatrix} \frac{\boldsymbol{p}_{nk}}{2} \\ u_k \end{pmatrix} \succ_{\bar{K}} 0$. Then strong duality holds. It is clear to show $\bar{K}^* = \left\{ \begin{pmatrix} \boldsymbol{x} \\ \beta_k \end{pmatrix} \in \mathbb{R}^{I+1} | \|\boldsymbol{x}\|_{m^*} \leq \beta_k \right\}$.

Using Lemma EC.2, $Z_1^{nk} \geq -\alpha_{nk} - v$ can be reformulated as the following constraints:

$$\begin{aligned} &-\boldsymbol{r}'_{nk}\boldsymbol{x} \geq -\alpha_{nk} - v, \forall n \in [N_k], k \in [K], \\ &\beta_k \geq \|x\|_{m^*}, \forall k \in [K], \end{aligned}$$

Following a similar approach, we can convert the remaining constraints into a sequence of finite constraints. $Z_2^{nk} := \inf_{\boldsymbol{r} \in \mathbb{R}^I, \rho(\boldsymbol{r}, \boldsymbol{r}_{nk}) \leq u_k} \alpha_{nk} + \beta_k u_k \geq 0$ can be reformulated as:

$$\alpha_{nk} \geq 0, \forall n \in [N_k], k \in [K].$$

Combining all the reformulated constraint in (EC.3), we obtain the following LP:

$$\begin{aligned} \min_{v, \boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}} \quad & v + \tfrac{1}{(1-\eta)} \left[\textstyle\sum_{k \in [K]} \sum_{n \in [N_k]} \frac{w_k}{N_k} \alpha_{nk} + \sum_{k \in [K]} \beta_k w_k \theta_k \right] \\ \text{s.t.} \quad & \boldsymbol{r}'_{nk}\boldsymbol{x} \geq -\alpha_{nk} - v, \forall n \in [N_k], k \in [K], \\ & \beta_k \geq \|x\|_{m^*}, \forall k \in [K], \\ & \alpha_{nk} \geq 0, \forall n \in [N_k], k \in [K], \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{EC.7}$$

Note that in (EC.7), if there exists $\beta_i^* \neq \beta_j^*$ for some $i, j \in [K]$ in the optimal solution, then we can always reduce both $\beta_i, \beta_j := \min(\beta_i^*, \beta_j^*)$ and keeping other variables the same to obtain a new feasible solution while reducing its objective value since $w_k\theta_k > 0, \forall k \in [K]$. Therefore, we have an implicit constraint $\beta_k = \beta, \forall k \in [K]$. Combining it with the current constraint, we obtain the reformulation, i.e. (12).

Then we show its computational tractability in some cases (i.e. $m = 1$ and $m = 2$). When $m = 1$, the second set of constraints in (12) is equivalent to:

$$\beta \geq x_i, \forall i \in [I],$$

$$\beta \geq -x_i, \forall i \in [I].$$

When $m = 2$, the second set of constraints in (12) is equivalent to:

$$(\beta, \boldsymbol{x}) \in \Omega_{soc}. \quad \square$$

Proof of Proposition EC.1. We have observed that the dual formula of the objective function can be eventually formulated as a linear program. We apply the same technique to handle the target constraint for the worst-case expected return. Consider

$$\begin{aligned} \inf_{\mathbb{P} \in \mathcal{U}_{rs}(\boldsymbol{\theta})} \mathbb{E}^{\mathbb{P}}[\boldsymbol{r}'\boldsymbol{x}] = \inf_{(\boldsymbol{r}, u_k) \sim \tilde{\mathbb{P}}_{nk}} \quad & \textstyle\sum_{k \in [K]} \sum_{n \in [N_k]} \frac{w_k}{N_k} \mathbb{E}^{\tilde{\mathbb{P}}_{nk}}[\boldsymbol{r}'\boldsymbol{x}] \\ \text{s.t.} \quad & \textstyle\sum_{n \in [N_k]} \frac{w_k}{N_k} \mathbb{E}^{\tilde{\mathbb{P}}_{nk}}[u_k] \leq w_k\theta_k, \forall k \in [K], \\ & \tilde{\mathbb{P}}_{nk}[\boldsymbol{r} \in \mathbb{R}^I, \rho(\boldsymbol{r}, \boldsymbol{r}_{nk}) \leq u_k] = 1, \forall n \in [N_k], k \in [K]. \end{aligned}$$

Denote the dual variables of each constraint as $\nu_k \ \forall k \in [K]$ and $\hat{\eta}_{nk} \ \forall n \in [N_k], k \in [K]$ respectively. We derive the dual problem under the proposed regime-switching ambiguity set as follow:

$$\begin{aligned}
&\sup_{\boldsymbol{\nu},\hat{\boldsymbol{\eta}}} \ \textstyle\sum_{k\in[K]}\sum_{n\in[N_k]} \hat{\eta}_{nk} - \sum_{k\in[K]} \nu_k w_k \theta_k \\
&\quad \text{s.t. } \hat{\eta}_{nk} - \tfrac{w_k}{N_k} u_k \nu_k \le \tfrac{w_k}{N_k} \boldsymbol{r'x}, \forall n \in [N_k], k \in [K], \boldsymbol{r} \in \mathbb{R}^I, \rho(\boldsymbol{r}, \boldsymbol{r}_{nk}) \le u_k, \\
&\quad \nu_k \ge 0, \forall k \in [K] \\
= &\sup_{\boldsymbol{\nu},\boldsymbol{\eta}} \ \textstyle\sum_{k\in[K]}\sum_{n\in[N_k]} \tfrac{w_k}{N_k}\eta_{nk} - \sum_{k\in[K]} \nu_k w_k \theta_k \\
&\quad \text{s.t. } \eta_{nk} - u_k \nu_k \le \boldsymbol{r'x}, \forall n \in [N_k], k \in [K], \boldsymbol{r} \in \mathbb{R}^I, \rho(\boldsymbol{r}, \boldsymbol{r}_{nk}) \le u_k, \\
&\quad \nu_k \ge 0, \forall k \in [K].
\end{aligned}$$

where $\eta_{nk} = \frac{N_k}{w_k}\hat{\eta}_{nk}$. There are infinite constraints. Following a similar analysis in Theorem 1, for $m \in [1, +\infty]$,

$$\begin{aligned}
\inf_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})} \mathbb{E}^{\mathbb{P}}[\boldsymbol{r'x}] = &\sup_{\nu,\boldsymbol{\eta}} \ \textstyle\sum_{k\in[K]}\sum_{n\in[N_k]} \tfrac{w_k}{N_k}\eta_{nk} - \nu \sum_{k\in[K]} w_k\theta_k \\
&\quad \text{s.t. } \boldsymbol{r}'_{nk}\boldsymbol{x} \ge \eta_{nk}, \forall n \in [N_k], k \in [K], \\
&\quad \nu \ge \|\boldsymbol{x}\|_{m^*}, \\
&\quad \nu \ge 0.
\end{aligned}$$

We can change the second constraint above similarly to reformulate as a set of linear and second order cone constraint(s) when $m = 1$ and $m = 2$ respectively. □

Proof of Proposition 2. The proof is similar to what has been shown in Pflug et al. (2012), where they prove the case when $K = 1$. Denote the center of Wasserstein ambiguity set as $\hat{\mathbb{P}} \in \mathcal{U}_{rs}(\mathbf{0})$. We rewrite the objective of our function (11) to be $\min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})} \text{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x})$ as CVaR metric implicitly minimize $v \in \mathbb{R}$. Given the final objective of Theorem 1, for any fixed portfolio weight $\boldsymbol{x}^0$, we have:

$$\sup_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})} \text{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x}^0) = \text{CVaR}_\eta^{\hat{\mathbb{P}}}(\boldsymbol{x}^0) + \frac{1}{1-\eta}\|\boldsymbol{x}^0\|_{m^*} \sum_{k\in[K]} w_k\theta_k, \tag{EC.8}$$

where m^* is the conjugate norm of our proposed ℓ_m-norm used in the Wasserstein distance.

Equation (EC.8) can be obtained by letting the set of portfolio weight to have only one element, i.e., $\mathcal{X} = \{\mathbf{x^0}\}$ in Theorem 1. Then $\beta^* = \|\boldsymbol{x}^0\|_{m^*}$ and $\alpha^*_{nk} = (-\boldsymbol{r}'_{nk}\boldsymbol{x}^0 - v)^+$ in the optimal solution of (12) and by definition of CVaR under $\hat{\mathbb{P}}$, we have:

$$\text{CVaR}_\eta^{\hat{\mathbb{P}}}(\boldsymbol{x}^0) = v + \frac{1}{1-\eta}\sum_{k\in[K]}\sum_{n\in[N_k]} \frac{w_k}{N_k}(-\boldsymbol{r}'_{nk}\mathbf{x^0} - v)^+,$$

which matches the right-hand side of (EC.8).

Then for the EW portfolio $\boldsymbol{x}^u = (\frac{1}{I}, ..., \frac{1}{I})'$, we would obtain a sufficient condition for the optimality of the EW portfolio (i.e. $\sup_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})} \text{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x}^u) \le \sup_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})} \text{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathcal{X}$) by some arithmetic manipulations:

$$\sum_{k\in[K]} w_k\theta_k \ge \frac{\|\boldsymbol{x} - \boldsymbol{x}^u\|_{m^*}}{\|\boldsymbol{x}\|_{m^*} - \|\boldsymbol{x}^u\|_{m^*}} \mathbb{E}^{\hat{\mathbb{P}}}[\|\boldsymbol{r}\|_m], \forall \boldsymbol{x} \in \mathcal{X}.$$

Therefore, as $\sum_{k\in[K]} w_k\theta_k \to \infty$, this sufficient condition would finally hold. Especially like (Pflug et al. 2012, Proposition 3), the solution of our DRO problem would converge to EW portfolio under ℓ_1-norm if $\sum_{k\in[K]} w_k\theta_k \ge (I-1)\mathbb{E}^{\hat{\mathbb{P}}}[\|\boldsymbol{r}\|_1]$. □

EC.2. Proof in Section 4

Before showing the proof, we first state the following several lemmas and definitions.

EC.2.1. Additional Definitions of Markov Chain

We follow the same definition in Assumption 1 and denote some other definitions extracted from Wolfer and Kontorovich (2021). Suppose the underlying true transition matrix of this Markov Chain is A^*.

Definition EC.1 (Spectral Gap). The *reversible* transition matrix A^* satisfies *detailed balance*, i.e. $\pi_i A^*_{i,j} = \pi_j A^*_{j,i}, \forall i,j \in [K]$ with π being the stationary distribution. Then for the eigenvalues of A^*: $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K$, the spectral gap is defined by:

$$\gamma_{\mathsf{s}} = 1 - \lambda_2(A^*).$$

Definition EC.2 (Pseudo-spectral Gap). Pseudo-spectral gap is defined by:

$$\gamma_{\mathsf{ps}} := \max_{m \geq 1} \left\{ \frac{\gamma_s((A^+)^m (A^*)^m)}{m} \right\},$$

where A^+ is the time reversal of A, given by $A^+_{i,j} := \frac{\pi_j A^*_{j,i}}{\pi_i}, \forall i,j \in [K]$.

EC.2.2. Technical Lemmas

Lemma EC.3 **(Estimation Error of Transition Matrix of Markov Chain)**. *Under Assumption 1, with probability $1-\delta$, as long as the window size N is large enough ($N \geq \frac{C}{\gamma_{ps}\pi_*}\log(\frac{K}{\delta}\sqrt{\frac{1}{\pi_*}})$), the output of the frequency-based approach w_i satisfies:*

$$|w_i - w_i^*| \leq \sqrt{\frac{CK\log(K/\delta)}{N\pi_*}}, \forall i \in [K], \tag{EC.9}$$

where C is the same constant in Theorem 3.1 in Wolfer and Kontorovich (2021).

Proof of Lemma EC.3. Recall $\pi_* = \min_{i\in[K]} \pi_i$ in Assumption 1, this result is a variant of Theorem 3.1 in Wolfer and Kontorovich (2021). The empirical estimator used in the proof of Theorem 3.1 in Wolfer and Kontorovich (2021) is exactly the estimation procedure of the probability transition matrix A shown in Section 4.1.1 as well as Step 3(a) in Algorithm 1. Therefore, compared with the underlying true transition matrix $A^* = \{a^*_{j,k}\}_{j,k\in[K]}$ of regime (state) sequences in Assumption 1, since the sample size $N \geq \frac{C}{\gamma_{\mathsf{ps}}\pi_*}\log(\frac{K}{\delta}\sqrt{\frac{1}{\pi_*}})$, then whenever $N \geq C\frac{K\log(K/\delta)}{\varepsilon^2\pi_*}$, we have with probability at least $1-\delta$, $\|A - A^*\|_\infty \leq \varepsilon$, which implies that $\|A - A^*\|_\infty \leq \sqrt{C}\sqrt{\frac{K\log(K/\delta)}{N\pi_*}}$ under large N with probability at least $1-\delta$. Finally, by definition of matrix norm, $\|A - A^*\|_\infty = \max_j \sum_{k\in[K]} |a_{j,k} - a^*_{j,k}|$ and $w_j = a_{i^*,j}$ for the last observed regime i^*, we can obtain the bound in (EC.9). □

Lemma EC.4 **(Empirical State Frequency Concentration)**. *Let $\{s_i\}_{i\geq 1}$ be a reversible Markov chain with stationary distribution π and spectral gap $\gamma_s > 0$. Then denote $N_k = \sum_{i=1}^N \mathbb{I}_{\{s_i = k\}}$, we have:*

$$\Pr\left(\left|\frac{N_k}{N} - \pi_k\right| \geq \varepsilon\right) \leq 2\exp(-2c_0' N\varepsilon^2), \tag{EC.10}$$

where c_0' is the same as defined in Proposition 3.

Proof of Lemma EC.4. This result is a direct application from Theorem 2.1 in Fan et al. (2021) by letting the bounded function $f_i(s) = \mathbb{I}_{\{s=k\}}, \forall i \in [N]$ in their condition. Then $\mathbb{E}^\pi[f_i(s)] = N\pi_s, \forall i \in [N]; \sum_{i=1}^N f_i(s_i) = N_k$ so that we first obtain from Theorem 2.1 in Fan et al. (2021), $\forall \varepsilon' > 0$:

$$\Pr(|N_k - N\pi_k| \geq \varepsilon') \leq 2\exp\left(-\frac{2c_0'\varepsilon'^2}{N}\right). \tag{EC.11}$$

Letting $\varepsilon = N\varepsilon'$, we would obtain the result of (EC.10). □

Lemma EC.5 **(Some Properties of Wasserstein Distance)**. *We need the following properties of Wasserstein Distance in our analysis:*

- *(a): Convexity (Lemma 2.10 in Pflug and Pichler (2014)):*

$$W_m(\mathbb{P},(1-\lambda)\mathbb{P}_0+\lambda\mathbb{P}_1)\le(1-\lambda)W_m(\mathbb{P},\mathbb{P}_0)+\lambda W_m(\mathbb{P},\mathbb{P}_1),\forall\lambda\in[0,1].$$

- *(b): Measure Concentration (Theorem 3.4 in Esfahani and Kuhn (2018)): if $\mathbb{P}\in\mathcal{P}(\mathbb{R}^d)(d\ge 2)$ is a light-tailed distribution (i.e. satisfying Assumption 3), ξ_i i.i.d. each sampling from $\mathbb{P}$ and $\hat{\mathbb{P}}_n=\frac{1}{n}\sum_{i\in[n]}\delta_{\xi_i}$, then $\forall t\in[0,1]$:*

$$\Pr(W_m(\mathbb{P},\hat{\mathbb{P}}_n)\ge t)\le c_1\exp(-c_2nt^d),$$

where c_1,c_2 are positive constants only depending on $\mathbb{P}$.

- *(c): Variational Representation (Theorem 3.2 in Esfahani and Kuhn (2018)): for any distribution $\mathbb{P}_1,\mathbb{P}_2\in\mathcal{P}(\mathbb{R}^d)$,*

$$W_m(\mathbb{P}_1,\mathbb{P}_2)=\sup_{f\in\mathcal{L}_m}\left(\mathbb{E}^{\mathbb{P}_1}[f(\xi)]-\mathbb{E}^{\mathbb{P}_2}[f(\xi)]\right),$$

where $\mathcal{L}_m$ denotes the space of all Lipschitz functions with $|f(\xi_1)-f(\xi_2)|\le\|\xi_1-\xi_2\|_m,\forall\xi_1,\xi_2\in\mathbb{R}^d$.

Lemma EC.6 **(Feasibility of Transportation Problem)**. *The following linear systems with variables $\{\mu_{ik}\}_{i\in[I],k\in[K]}$ is feasible if $\{w_i^*\}_{i\in[I]}$ and $\{w_k\}_{k\in[K]}$ are all nonnegative and $\sum_{i\in[I]}w_i^*=\sum_{k\in[K]}w_k$:*

$$\begin{aligned}\sum_{k\in[K]}\mu_{ik}&=w_i^*,\ \forall i\in[I];\\ \sum_{i\in[I]}\mu_{ik}&=w_k,\ \forall k\in[K];\\ \mu_{ik}&\ge 0,\ \forall i\in[I],k\in[K].\end{aligned}$$

This lemma follows directly from Chapter 6 in Luenberger et al. (1984) since it can be regarded as the feasible region of one balanced transportation problem.

The following lemma serves as a building block to show Theorem 2, which decomposes $\mathbb{P}^*$ with a mixture of "proper distributions" $\hat{\mathbb{P}}_k^*$ with their regime frequencies by $\{w_k\}_{k\in[K]}$ instead of the true $\{w_k^*\}_{k\in[K]}$.

Lemma EC.7 **(Probability Reconstruction)**. *We can represent the underlying true distribution $\mathbb{P}^*$ to be $\mathbb{P}^*=\sum_{k\in[K]}w_k\hat{\mathbb{P}}_k^*$, where $\hat{\mathbb{P}}_k^*=\sum_{i\in[K]}\alpha_{ik}\mathbb{P}_i^*,\forall k\in[K]$ with a set of nonnegative $\{\alpha_{ik}\}_{i,k\in[K]}$ and $\sum_{i\in[K]}\alpha_{ik}=1,\forall k\in[K]$. And $\forall k\in[K]$,*

$$\alpha_{kk}=\begin{cases}1, & \text{if } w_k\le w_k^*\\ \frac{w_k^*}{w_k}, & \text{if } w_k>w_k^*\end{cases}. \tag{EC.12}$$

Proof of Lemma EC.7. Since $\mathbb{P}^*=\sum_{k\in[K]}w_k^*\mathbb{P}_k^*$ from Assumption 2, if we want $\mathbb{P}^**=\sum_{k\in[K]}w_k\hat{\mathbb{P}}_k^*$, we need to show that we can represent $w_k^*=\sum_{i\in[K]}\alpha_{ki}w_i,\forall k\in[K]$. It is equivalent to say that α_{ik} need to satisfy the following constraints:

$$\begin{aligned}\sum_{k\in[K]}\alpha_{ik}w_k&=w_i^*,\ \forall i\in[K];\\ \sum_{i\in[K]}\alpha_{ik}&=1,\ \forall k\in[K];\\ \alpha_{ik}&\ge 0,\ \forall i,k\in[K].\end{aligned} \tag{EC.13}$$

where the second and third set of constraints is to guarantee that the new constructed $\hat{\mathbb{P}}_k^* := \sum_{i\in[K]} \alpha_{ik}\mathbb{P}_i^*$ is a distribution. We need to show (EC.13) is feasible for $\{\alpha_{ik}\}_{i,k\in[K]}$ satisfying (EC.12). That is to say, we need to find a set of feasible solution of $\{\alpha_{ik}\}_{i,k\in[K]}$ in (EC.13) satisfying (EC.12).

Denote $\tilde{K} = \{\tilde{k} | w_{\tilde{k}} = 0, \tilde{k} \in [K]\}$. If $\tilde{K} \neq \emptyset$, for those variables $\{\alpha_{i\tilde{k}}\}_{i\in[K],\tilde{k}\in\tilde{K}}$, we assign $\alpha_{i\tilde{k}} = \begin{cases} 1, \text{ if } i = \tilde{k} \\ 0, \text{ otherwise} \end{cases}$ such that the assigned $\{\alpha_{ik}\}_{i\in[K],k\in\tilde{K}}$ satisfy (EC.12) and feasible in (EC.13) for the constraints they appear. Then we need to find one feasible $\{\alpha_{ik}\}_{i\in[K],k\in[K]\setminus\tilde{K}}$ to satisfy (EC.12) and (EC.13) under the value assignments of $\{\alpha_{ik}\}_{i\in[K],k\in\tilde{K}}$.

Since $w_k > 0$ for $k \in [K]\setminus\tilde{K}$, if we define $\mu_{ik} = w_k\alpha_{ik}, \forall i,k \in [K]\setminus\tilde{K}$, then the feasibility of $\{\alpha_{ik}\}_{i\in[K],k\in[K]\setminus\tilde{K}}$ in (EC.13) is equivalent to that of $\{\mu_{ik}\}_{i\in[K],k\in[K]\setminus\tilde{K}}$ due to one to one correspondence with the following constraints:

$$\begin{aligned} \sum_{k\in[K]\setminus\tilde{K}} \mu_{ik} = w_i^*,\ \forall i \in [K]; \\ \sum_{i\in[K]} \mu_{ik} = w_k,\ \forall k \in [K]\setminus\tilde{K}; \\ \mu_{ik} \geq 0,\ \forall i \in [K], k \in [K]\setminus\tilde{K}. \end{aligned} \tag{EC.14}$$

Based on the discussion above, it is equivalent to showing that there is a set of feasible solution $\{\mu_{ik}\}_{i\in[K],k\in[K]\setminus\tilde{K}}$ where $\mu_{kk} = \min\{w_k, w_k^*\}, \forall k \in [K]\setminus\tilde{K}$ in (EC.14).

Define $I_1 = \{i \in [K] : w_i \geq w_i^*\}, K_1 = \{k \in [K] : 0 < w_k < w_k^*\}$. Then $I_1 \cup K_1 = [K]\setminus\tilde{K}, I_1 \cap K_1 = \emptyset$. Since $\mu_{kk} = \min\{w_k, w_k^*\}, \forall k \in [K]\setminus\tilde{K}$, by (EC.14) we have:

- If $i \in I_1$, then $\mu_{ii} = w_i^*; \mu_{ik} = 0, \forall k \neq i, k \in [K]\setminus\tilde{K}$;
- If $k \in K_1$, then $\mu_{kk} = w_k; \mu_{ik} = 0, \forall i \neq k, i \in [K]$.

Then by computation, the remaining unassigned $\{\mu_{ik}\}_{i\in K_1\cup\tilde{K},k\in I_1}$ would need to satisfy following constraints under our assignments of other variables above:

$$\begin{aligned} \sum_{k\in I_1} \mu_{ik} = w_i^* - w_i,\ \forall i \in K_1 \cup \tilde{K}; \\ \sum_{i\in K_1\cup\tilde{K}} \mu_{ik} = w_k - w_k^*,\ \forall k \in I_1; \\ \mu_{ik} \geq 0,\ \forall i \in K_1 \cup \tilde{K}, k \in I_1. \end{aligned} \tag{EC.15}$$

And the region (EC.15) is feasible for $\{\mu_{ik}\}_{i\in K_1\cup\tilde{K},k\in I_1}$ since $\sum_{i\in K_1\cup\tilde{K}}(w_i^* - w_i) = \sum_{k\in I_1}(w_i - w_i^*) > 0$ by definition of $\tilde{K}, I_1, K_1$ from Lemma EC.6. Take one feasible solution $\{\mu_{ik}\}_{i\in K_1\cup\tilde{K},k\in I_1}$ and by previous assignments, we find a feasible solution for $\{\mu_{ik}\}_{i\in[K],k\in[K]\setminus\tilde{K}}$ satisfying (EC.14) with $\mu_{kk} = \min\{w_k, w_k^*\}, \forall k \in [K]\setminus\tilde{K}$.

Then for any feasible $\{\mu_{ik}\}_{i\in[K],k\in[K]\setminus\tilde{K}}$ above, it corresponds to one feasible $\{\alpha_{ik}\}_{i\in[K],k\in[K]\setminus\tilde{K}}$ under our assignment for $\{\alpha_{i\tilde{k}}\}_{i\in[K],\tilde{k}\in\tilde{K}}$. Therefore, we obtain a set of feasible solution of $\{\alpha_{ik}\}_{i,k\in[K]}$ in (EC.13) satisfying (EC.12) and show $\mathbb{P}^* = \sum_{k\in[K]} w_k(\sum_{i\in[K]} \alpha_{ik}\mathbb{P}_i^*)$ for some $\{\alpha_{ik}\}_{i,k\in[K]}$ satisfying (EC.12). □

EC.2.3. Proof of Main Results

Proof of Theorem 2. By Lemma EC.7, we can represent $\mathbb{P}^* = \sum_{k\in[K]} w_k\hat{\mathbb{P}}_k^*$ where $\hat{\mathbb{P}}_k^* = \sum_{i\in[K]} \alpha_{ik}\mathbb{P}_i^*, \forall k \in [K]$ with the same requirement for $\{\alpha_{ik}\}_{i,k\in[K]}$.

Then for each regime occurring with probability w_k, by the representation of our ambiguity set in Lemma EC.1, we only need to consider $\theta_k \geq W_m(\hat{\mathbb{P}}_k, \hat{\mathbb{P}}_k^*)$ to cover the true distribution. And by triangle inequality, we have:

$$W_m(\hat{\mathbb{P}}_k^*, \hat{\mathbb{P}}_k) \leq W_m(\mathbb{P}_k^*, \hat{\mathbb{P}}_k) + W_m(\mathbb{P}_k^*, \hat{\mathbb{P}}_k^*) \tag{EC.16}$$

We aim to bound the two terms of right-hand side with high probability.

For the first term of right-hand side in (EC.16), we leverage on the result of the measure concentration of Wasserstein distance (i.e. Lemma EC.5 (b)). Specially, $\hat{\mathbb{P}}_k$ is the empirical distribution from $\mathbb{P}_k^*$ and we have:

$$\Pr(W_m(\mathbb{P}_k^*,\hat{\mathbb{P}}_k)\ge t)\le \hat{c}_1\exp(-\hat{c}_2 N_k t^I), \tag{EC.17}$$

where $\hat{c}_1,\hat{c}_2$ are positive constants only depending on $\mathbb{P}_k^*$. Then equating the right-hand side of (EC.17) to $\frac{\delta}{2K}$ and solving for t yields $t=\left(\frac{\log(2K\hat{c}_1/\delta)}{\hat{c}_2 N_k}\right)^{\frac{1}{I}}$. Then letting $c_{k,1}=\frac{1}{\hat{c}_2}, c_{k,2}=\log(2\hat{c}_1)$, we would have: $W_m(\mathbb{P}_k^*,\hat{\mathbb{P}}_k)\le\left(\frac{c_{k,1}\log(K/\delta)+c_{k,2}}{N_k}\right)^{\frac{1}{I}}$ with probability at least $1-\frac{\delta}{2K}$. Therefore, denote the event $A_k=\left\{W_m(\mathbb{P}_k^*,\hat{\mathbb{P}}_k)\le\left(\frac{c_{k,1}\log(K/\delta)+c_{k,2}}{N_k}\right)^{\frac{1}{I}}\right\}$, we have $\Pr(A_k)\ge 1-\frac{\delta}{2K},\forall k\in[K]$.

For the second term of right-hand side in (EC.16), if $w_k=0$, then $\hat{\mathbb{P}}_k^*=\mathbb{P}_k^*$ by (EC.12) and the second term would disappear. Otherwise for $w_k\neq 0$, we have:

$$W_m(\mathbb{P}_k^*,\hat{\mathbb{P}}_k^*)=W_m(\mathbb{P}_k^*,\sum_{i\in[K]}\alpha_{ik}\mathbb{P}_i^*)\le\sum_{i\in[K],i\neq k}\alpha_{ik}W_m(\mathbb{P}_k^*,\mathbb{P}_i^*)\le(1-\min_{k\in[K]}\min\{\frac{w_k^*}{w_k},1\})\max_{i\in[K]}W_m(\mathbb{P}_i^*,\mathbb{P}_k^*), \tag{EC.18}$$

where the first inequality follows from Lemma EC.5(a). Specially by repeating using convexity property, we have:

$$\begin{aligned}
W_m(\mathbb{P}_k^*,\sum_{i\in[K]}\alpha_{ik}\mathbb{P}_i^*)&=W_m(\mathbb{P}_k^*,\alpha_{1k}\mathbb{P}_1^*+(1-\alpha_{1k})\sum_{i>1}\frac{\alpha_{ik}\mathbb{P}_i^*}{1-\alpha_{1k}})\\
&\le\alpha_{1k}W_m(\mathbb{P}_k^*,\mathbb{P}_1^*)+(1-\alpha_{1k})W_m(\mathbb{P}_k^*,\sum_{i>1}\frac{\alpha_{ik}\mathbb{P}_i^*}{1-\alpha_{1k}})\\
&\le\alpha_{1k}W_m(\mathbb{P}_k^*,\mathbb{P}_1^*)+\alpha_{2k}W_m(\mathbb{P}_k^*,\mathbb{P}_2^*)+(1-\alpha_{1k}-\alpha_{2k})W_m\left(\mathbb{P}_k^*,\sum_{i>2}\frac{\alpha_{ik}\mathbb{P}_i^*}{1-\alpha_{1k}-\alpha_{2k}}\right)\\
&\ldots\le\sum_{i\in[K]}\alpha_{ik}W_m(\mathbb{P}_k^*,\mathbb{P}_i^*)=\sum_{i\in[K],i\neq k}\alpha_{ik}W_m(\mathbb{P}_k^*,\mathbb{P}_i^*).
\end{aligned}$$

And the second inequality follows by noticing:

$$\begin{aligned}
\sum_{i\in[K],i\neq k}\alpha_{ik}W_m(\mathbb{P}_k^*,\mathbb{P}_i^*)&\le(\sum_{i\in[K],i\neq k}\alpha_{ik})\max_{i\in[K]}W_m(\mathbb{P}_i^*,\mathbb{P}_k^*)\\
&\le(1-\min\{\frac{w_k^*}{w_k},1\})\max_{i\in[K]}W_m(\mathbb{P}_i^*,\mathbb{P}_k^*),
\end{aligned}$$

where the last inequality follows by observing $\sum_{i\in[K],i\neq k}\alpha_{ik}=1-\alpha_{kk}=1-\min\{\frac{w_k^*}{w_k},1\}\le 1-\min_{k\in[K]}\min\{\frac{w_k^*}{w_k},1\}$.

Furthermore, we have:

$$\begin{aligned}
\min_{k\in[K]}\min\left\{\frac{w_k^*}{w_k},1\right\}&=1+\min_{k\in[K]}\min\left\{\frac{w_k^*-w_k}{w_k},0\right\}\\
&\ge 1-\max_{k\in[K]}\frac{|w_k^*-w_k|}{w_k}\ge 1-\frac{\max_{k\in[K]}|w_k-w_k^*|}{\min_{i\in[K],w_i\neq 0}w_i}
\end{aligned}$$

From Lemma EC.3, if the sample size $N\ge\frac{c_0^2}{\gamma_{\mathsf{ps}}\pi_*}\log(\frac{2K}{\delta}\sqrt{\frac{1}{\pi_*}})$, then with probability at least $1-\frac{\delta}{2}$, $|w_k-w_k^*|\le c_0\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}},\forall k\in[K]$, which implies that: $\min_{k\in[K]}\min\{\frac{w_k^*}{w_k},1\}\ge 1-$

$\frac{c_0}{\min_{i\in[K],w_i\neq 0} w_i}\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}}$. Therefore, denote the event $B=\left\{|w_k-w_k^*|\le c_0\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}}, \forall k\in[K]\right\}$. Then $\Pr(B)\ge 1-\frac{\delta}{2}$, and under the event B, we have: $\min_{k\in[K]}\min\{\frac{w_k^*}{w_k},1\}\ge 1-\frac{c_0}{\min_{i\in[K],w_i\neq 0} w_i}\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}}$.

Finally, $\underline{\text{under the event } (\bigcap_{k\in[K]} A_k)\cap B}$, following the decomposition in (EC.16) and (EC.18), we have:

$$W_m(\hat{\mathbb{P}}_k^*,\hat{\mathbb{P}}_k)\le\left(\frac{c_{k,1}\log(K/\delta)+c_{k,2}}{N_k}\right)^{\frac{1}{I}}+\frac{c_0}{\min_{i\in[K],w_i\neq 0} w_i}\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}}\max_{i\neq k} W_m(\mathbb{P}_k^*,\mathbb{P}_i^*), \forall k\in[K]. \tag{EC.19}$$

Therefore, if each component θ_k in $\boldsymbol{\theta}$ is larger than the right-hand side of (EC.19) for each $k\in[K]$, then $\theta_k\ge W_m(\hat{\mathbb{P}}_k^*,\hat{\mathbb{P}}_k)$. And due to $\mathbb{P}^*=\sum_{k\in[K]} w_k\hat{\mathbb{P}}_k^*$ and the representation $\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta})$ in Lemma EC.1, the true distribution would lie in the ambiguity set, i.e. $\mathbb{P}^*\in\Pi_{\tilde{\boldsymbol{r}}}\mathcal{U}_{rs}(\boldsymbol{\theta})$.

And we only need to show the event $(\bigcap_{k\in[K]} A_k)\cap B$ happen with probability at least $1-\delta$ to reach our conclusion. This is given by:

$$\begin{aligned}\Pr((\bigcap_{k\in[K]} A_k)\cap B) &= 1-\Pr(\bigcup_{k\in[K]} A_k^C\cup B^C)\\ &\ge 1-\sum_{k\in[K]}\Pr(A_k^C)-\Pr(B^C)\\ &\ge 1-K\times\frac{\delta}{2K}-\frac{\delta}{2}=1-\delta.\end{aligned}$$

□

Proof of Proposition 3. We follow the proof notation in Theorem 2.

For any $\delta>0$, under the event $B=\left\{|w_k-w_k^*|\le c_0\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}}, \forall k\in[K]\right\}$, if $N\ge\frac{c_0^2K\log(2K/\delta)}{\pi_*(\min_{i\in[K]} w_i^*)^2}$, then we have: $|w_k-w_k^*|\le\min_{i\in[K]} w_i^*\le w_k^*, \forall k\in[K]$. And $\underline{\min_{i\in[K]} w_i\le 2\min_{i\in[K]} w_i^*}$.

We also denote the event $C_k=\{|\frac{N_k}{N}-\pi_k|\le\sqrt{\frac{\log(2K/\delta)}{2c_0'N}}\}$ such that the empirical state frequency s close to the stationary distribution. $\Pr(C_K)\ge 1-\frac{\delta}{K}$ by applying Lemma EC.4 and equating the right-hand of (EC.10) to be $\frac{\delta}{K}$. Then if $\pi_k\ge\sqrt{\frac{\log(2K/\delta)}{2c_0'N}}$, i.e. $N\ge\frac{\log(2K/\delta)}{2c_0'\pi_k^2}$, then $|\frac{N_k}{N}-\pi_k|\le\pi_k$ such that $\underline{N_k\le 2N\pi_k}$ under event C_k.

Therefore, combining the argument under $(\bigcap_{k\in[K]} A_k)\cap B\cap(\bigcap_{k\in[K]} C_k)$ and the condition of N we have:

$$\begin{aligned}\theta_k &\ge\left(\frac{c_{k,1}\log(K/\delta)+c_{k,2}}{N_k}\right)^{\frac{1}{I}}+\frac{c_0}{\min_{i\in[K],w_i\neq 0} w_i}\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}}\max_{i\neq k} W_m(\mathbb{P}_k^*,\mathbb{P}_i^*)\\ &\ge\left(\frac{c_{k,1}\log(K/\delta)+c_{k,2}}{2N\pi_k}\right)^{\frac{1}{I}}+\frac{c_0}{2\min_{i\in[K]} w_i^*}\sqrt{\frac{K\log(2K/\delta)}{N\pi_*}}\max_{i\neq k} W_m(\mathbb{P}_k^*,\mathbb{P}_i^*).\end{aligned}$$

And $\Pr((\bigcap_{k\in[K]} A_k)\cap B\cap(\bigcap_{k\in[K]} C_k))\ge 1-2\delta$ by the same argument before. Then replacing δ with $\frac{\delta}{2}$ above would obtain the result in (15). □

Proof of Theorem 3. Denote $\tilde{G}((\boldsymbol{x},v),\boldsymbol{r}):=v+\frac{1}{1-\eta}G(\boldsymbol{x},\boldsymbol{r})$ in (6), then $\mathrm{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x})=\min_{v\in\mathbb{R}}\mathbb{E}^{\mathbb{P}}[\tilde{G}(\boldsymbol{x},v),\boldsymbol{r}]$. Suppose $(\boldsymbol{x}^*,v^*)\in\arg\min_{\boldsymbol{x},v}\mathbb{E}^{\mathbb{P}^*}[\tilde{G}((\boldsymbol{x},v),\boldsymbol{r})]$, then $\mathbb{E}^{\mathbb{P}^*}[\tilde{G}((\boldsymbol{x}^*,v^*),\boldsymbol{r})]=\mathrm{CVaR}_\eta^{\mathbb{P}^*}(\boldsymbol{x}^*)$ by definition. Similarly, we define:

$$(\hat{\boldsymbol{x}},\hat{v})\in\arg\min_{\boldsymbol{x},v}\sup_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathbb{E}^{\mathbb{P}}[\tilde{G}((\boldsymbol{x},v),\boldsymbol{r})].$$

Notice that $\forall \boldsymbol{x} \in \mathcal{X}, v \in \mathbb{R}$, we have:

$$|\tilde{G}((\boldsymbol{x},v),\boldsymbol{r}_1) - \tilde{G}((\boldsymbol{x},v),\boldsymbol{r}_2)| \leq |\frac{1}{1-\eta}\boldsymbol{x}'(\boldsymbol{r}_1-\boldsymbol{r}_2)| \leq \frac{\|\boldsymbol{x}\|_{m^*}}{1-\eta}\|\boldsymbol{r}_1-\boldsymbol{r}_2\|_m, \forall m \geq 1.$$

To prove the first inequality, we note that $|\tilde{G}((\boldsymbol{x},v),\boldsymbol{r}_1) - \tilde{G}((\boldsymbol{x},v),\boldsymbol{r}_2)| = |\frac{1}{1-\eta}G(\boldsymbol{x},\boldsymbol{r}_1) - \frac{1}{1-\eta}G(\boldsymbol{x},\boldsymbol{r}_2)|$. Recall from the definition of $G(\boldsymbol{x},\boldsymbol{r}_1) = (-\boldsymbol{x}'\boldsymbol{r}_1 - v)^+$, we can prove the inequality by considering the following three cases:

(i) if $-\boldsymbol{x}'\boldsymbol{r}_1 - v \geq 0$ and $-\boldsymbol{x}'\boldsymbol{r}_2 - v \leq 0$: $|G(\boldsymbol{x},\boldsymbol{r}_1) - G(\boldsymbol{x},\boldsymbol{r}_2)| = -\boldsymbol{x}'\boldsymbol{r}_1 - v \leq -\boldsymbol{x}'\boldsymbol{r}_1 + \boldsymbol{x}'\boldsymbol{r}_2 \leq |\boldsymbol{x}'(\boldsymbol{r}_1-\boldsymbol{r}_2)|$.

(ii) if $-\boldsymbol{x}'\boldsymbol{r}_1 - v \leq 0$ and $-\boldsymbol{x}'\boldsymbol{r}_2 - v \geq 0$: $|G(\boldsymbol{x},\boldsymbol{r}_1) - G(\boldsymbol{x},\boldsymbol{r}_2)| = -\boldsymbol{x}'\boldsymbol{r}_2 - v \leq -\boldsymbol{x}'\boldsymbol{r}_2 + \boldsymbol{x}'\boldsymbol{r}_1 \leq |\boldsymbol{x}'(\boldsymbol{r}_1-\boldsymbol{r}_2)|$.

(iii) if $-\boldsymbol{x}'\boldsymbol{r}_1 - v$ and $-\boldsymbol{x}'\boldsymbol{r}_2 - v$ are of the same sign: $|G(\boldsymbol{x},\boldsymbol{r}_1) - G(\boldsymbol{x},\boldsymbol{r}_2)| = |\boldsymbol{x}'(\boldsymbol{r}_1-\boldsymbol{r}_2)|$.

The second inequality is from the definition of dual norm. Therefore, $\tilde{G}((\boldsymbol{x},v),\boldsymbol{r})$ is Lipschitz continuous with respect to $\boldsymbol{r}$ with constant upper bounded by $\frac{\|\boldsymbol{x}\|_m^*}{1-\eta}$. Based on Lemma EC.5(c), we have:

$$\mathbb{E}^{\hat{\mathbb{P}}_k^*}[\tilde{G}((\boldsymbol{x},v),\boldsymbol{r})] - \mathbb{E}^{\mathbb{P}_k^*}[\tilde{G}((\boldsymbol{x},v),\boldsymbol{r})] \leq \frac{\|\boldsymbol{x}\|_{m^*}}{1-\eta}W_m(\hat{\mathbb{P}}_k^*,\mathbb{P}_k^*), \forall k \in [K], \forall x \in \mathcal{X}, v \in \mathbb{R}.$$

Now we can decompose the generalization error of our DRO model as follows:

$$\begin{aligned}
&\mathrm{CVaR}_\eta^{\mathbb{P}^*}(\hat{\boldsymbol{x}}) - \mathrm{CVaR}_\eta^{\mathbb{P}^*}(\boldsymbol{x}^*) = (\mathrm{CVaR}_\eta^{\mathbb{P}^*}(\hat{\boldsymbol{x}}) - \sup\nolimits_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathrm{CVaR}_\eta^{\mathbb{P}}(\hat{\boldsymbol{x}}))\\
&+(\sup\nolimits_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathrm{CVaR}_\eta^{\mathbb{P}}(\hat{\boldsymbol{x}}) - \sup\nolimits_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathrm{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x}^*)) + (\sup\nolimits_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathrm{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x}^*) - \mathrm{CVaR}_\eta^{\mathbb{P}^*}(\boldsymbol{x}^*))\\
&\leq \sup\nolimits_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathrm{CVaR}_\eta^{\mathbb{P}}(\boldsymbol{x}^*) - \mathrm{CVaR}_\eta^{\mathbb{P}^*}(\boldsymbol{x}^*)\\
&\leq \sup\nolimits_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathbb{E}^{\mathbb{P}}[\tilde{G}((\boldsymbol{x}^*,v^*),\boldsymbol{r})] - \mathbb{E}^{\mathbb{P}^*}[\tilde{G}((\boldsymbol{x}^*,v^*),\boldsymbol{r})]\\
&= \sum\nolimits_{k\in[K]} w_k(\sup\nolimits_{\mathbb{P}_k\in\mathcal{U}_{rs}^k(\theta_k)}\mathbb{E}^{\mathbb{P}_k}[\tilde{G}((\boldsymbol{x}^*,v^*),\boldsymbol{r})] - \mathbb{E}^{\hat{\mathbb{P}}_k^*}[\tilde{G}((\boldsymbol{x}^*,v^*),\boldsymbol{r})]).\\
&\leq \tfrac{\|\boldsymbol{x}^*\|_{m^*}}{1-\eta}\sum\nolimits_{k\in[K]} w_k \sup\nolimits_{\mathbb{P}_k\in\mathcal{U}_{rs}^k(\theta_k)} W_m(\mathbb{P}_k,\hat{\mathbb{P}}_k^*)\\
&\leq \tfrac{\|\boldsymbol{x}^*\|_{m^*}}{1-\eta}\sum\nolimits_{k\in[K]} w_k \left(\sup\nolimits_{\mathbb{P}_k\in\mathcal{U}_{rs}^k(\theta_k)} W_m(\mathbb{P}_k,\hat{\mathbb{P}}_k) + W_m(\hat{\mathbb{P}}_k,\hat{\mathbb{P}}_k^*)\right) \leq \tfrac{2\|\boldsymbol{x}^*\|_{m^*}}{1-\eta}\sum\nolimits_{k\in[K]} w_k\theta_k.
\end{aligned}$$

The first inequality follows from the fact that $\mathbb{P}^* \in \mathcal{U}_{rs}(\boldsymbol{\theta})$ with probability $1-\delta$ and the optimum of $\hat{\boldsymbol{x}}$, which reduces the first two terms to non-positive. And the second equality follows from the contained distribution in the ambiguity set $\mathcal{U}_{rs}(\boldsymbol{\theta}) = \{\mathbb{P} : \mathbb{P} = \sum_{k\in[K]} w_k\mathbb{P}_k, \mathbb{P}_k \in \mathcal{U}_{rs}^k(\theta_k), \forall k \in [K]\}$ and $\mathbb{P}^* = \sum_{k\in[K]} w_k\hat{\mathbb{P}}_k^*$. The fourth inequality is from the triangle inequality. The fifth inequality follows from $\hat{\mathbb{P}}_k^* \in \mathcal{U}_{rs}^k(\theta_k)$. $\square$

EC.3. Distributionally Robust Mean-CVar Model

In the robust Mean-CVaR model, the target is on the *worst-case* expected return of a portfolio; see Kang et al. (2019). In our application, a distributionally robust mean-CVaR portfolio optimization under regime-switching ambiguity set solves the constrained problem (11) with an additional first-moment constraint:

$$\min_{\boldsymbol{x}\in\mathcal{X}, v\in\mathbb{R}} \left\{ v + \frac{1}{1-\eta}\sup_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathbb{E}^{\mathbb{P}}[G(\boldsymbol{x},\boldsymbol{r})] \right\} \quad \text{s.t.} \quad \inf_{\mathbb{P}\in\mathcal{U}_{rs}(\boldsymbol{\theta})}\mathbb{E}^{\mathbb{P}}[\boldsymbol{r}'\boldsymbol{x}] \geq R. \tag{EC.20}$$

The following proposition shows that even with worst-case mean constraint, the robust portfolio optimization problem is still tractable. The additional worst-case mean constraint brings an extra set of constraints to the problem (11).

Proposition EC.1. *$\forall m \in [1, +\infty]$, the problem* (EC.20) *can be formulated to be the problem* (12) *with some additional constraints, given as follows:*

$$\begin{aligned}
&\textstyle\sum_{k\in[K]}\sum_{n\in[N_k]}\frac{w_k}{N_k}\eta_{nk} - \nu\sum_{k\in[K]} w_k\theta_k \geq R,\\
&\boldsymbol{r}'_{nk}\boldsymbol{x} \geq \eta_{nk}, \forall n \in [N_k], k \in [K],\\
&\nu \geq \|x\|_{m^*},\\
&\nu \geq 0,
\end{aligned} \tag{EC.21}$$

where ν, η_{nk} $\forall n \in [N_k], k \in [K]$ are additional dual variables introduced.

In particular, when $m = 1$ and $m = 2$, (EC.21) *is a set of linear and second-order conic constraints, respectively, which is tractable.*

Similarly, if we use the Wasserstein metric with ℓ_2-norm, the corresponding reformulation is a second-conic program. In our experiments, we will numerically examine both (robust) CVaR and Mean-CVaR portfolios in ℓ_1-norm.

EC.4. Detailed Cross validation algorithm

Algorithm 4 Specify the ambiguity set size through cross-validation

```
Input: Return sequences $\mathcal{R}_t$, set of candidate aversion level $\Theta$, number of folds $m$ (assume $\frac{N}{m} \in \mathbb{Z}$), associated
    observed regime sequences $\mathcal{S}_t := \{s_t, \dots, s_{t+N-1}\}$ (frequency-based) or $\mathcal{O}_t = \{o_t, \dots, o_{t+N-1}\}$ (HMM).
Partition $\mathcal{R}_t$ into $m$ equally sized subsets with $\mathcal{R}_{t,i} = \{\boldsymbol{r}_{t+\frac{N(i-1)}{m}}, \dots, \boldsymbol{r}_{t+\frac{Ni}{m}-1}\}, i \in [m]$.
for $i = 2, \dots, m$ do
    for $\boldsymbol{\theta} \in \Theta$ do
        for $j = 0, \dots, \frac{N}{m} - 1$ do
            Obtain $\{w_k, \{\boldsymbol{r}_{nk}\}_{n\in[N_k]}\}_{k\in[K]}$ following the modified Algorithm 2 (input with $\mathcal{D}_t \backslash \mathcal{R}_{t,i}, \mathcal{S}_t \backslash \mathcal{S}_{t,i}$) or Algo-
            rithm 3 (input with $\mathcal{R}_t \backslash \mathcal{R}_{t,i}, \mathcal{O}_t \backslash \mathcal{O}_{t,i}$), where the only modification is to set the last regime $j^*$ to
            $s_{t+\frac{N(i-1)}{m}+j-1}$ in Step 2 of Algorithm 2 or 3.
            Solve the optimization problem in (12) and obtain $x^*(\boldsymbol{\theta}; \mathcal{R}_t \backslash \mathcal{R}_{t,i}, s_{t+\frac{N(i-1)}{m}+j-1})$.
        end for
    end for
    Find the best parameter $\boldsymbol{\theta} \in \Theta$ that yields the highest out-of-sample Sharpe ratio (define in Appendix EC.5)
    on the validation dataset $\mathcal{R}_{t,i}$:
        $$\boldsymbol{\theta}^*_{-i} \in \arg\max_{\boldsymbol{\theta}\in\Theta} \text{ SharpeRatio}\left(\{\boldsymbol{r}'_{t+\frac{N(i-1)}{m}+j}\boldsymbol{x}^*(\boldsymbol{\theta}; \mathcal{R}_t \backslash \mathcal{R}_{t,i}, s_{t+\frac{N(i-1)}{m}+j-1})\}_{j=0}^{N/m-1}\right).$$
end for
Output: optimal $\boldsymbol{\theta}^* = \frac{1}{m-1}\sum_{i=2}^{m}\boldsymbol{\theta}^*_{-i}$.
```

EC.5. Performance Metrics

We set $M = T$ as the time span in the entire dataset and use $N = 120$ and 36 for the DeMiguel et al. (2009)'s datasets and our own dataset, respectively in Algorithm 1. Given the out-of-sample returns, we compute four performance metrics for each portfolio as follows. First, we calculate the out-of-sample Sharpe ratio, defined as the sample mean of the out-of-sample excess returns over the risk-free asset , $\hat{\mu} = \frac{1}{M-N}\sum_{t=1}^{M-N}\hat{g}_{N+t}$, then divided by their sample standard deviation, $\hat{\sigma} = \sqrt{\frac{1}{M-N-1}\sum_{t=1}^{M-N}(\hat{g}_{N+t} - \hat{\mu})^2}$:

$$\text{Sharpe ratio} = \frac{\hat{\mu}}{\hat{\sigma}}.$$

Second, we measure the certainty equivalent (CEQ) return, defined as the risk-free rate that an investor is willing to take compared to using a particular risky strategy. The formula is defined as:

$$\text{CEQ} = \hat{\mu} - \frac{1}{2}\hat{\sigma}^2.$$

Third, we use the maximum drawdown to measure the downside risk of a portfolio over a specific time period. This metric is computed as the maximum observed relative loss from a peak to a trough before attaining a new peak for a given portfolio. Specifically, we use $\{V_k\}_{k=0}^{M-N}$ to denote the accumulated wealth series of the portfolio from the beginning with unit 1, i.e., $V_k = \begin{cases} \prod_{t=1}^{k}(1+\hat{g}_{N+t}), \text{ if } k \geq 1, \\ 1, \text{ if } k = 0 \end{cases}$. We define the return peak indices to be $\mathcal{N}_p = \{j \geq 1 | V_j \geq \max\{V_{j-1}, V_{j+1}\}\}$ and denote the element in $\mathcal{N}_p$ as $\{N_p\}_{p\in[|\mathcal{N}_p|+1]}$ with $N_{p+1} = +\infty$. Then, the maximum drawdown is given by:

$$\text{Maximum Drawdown} = 1 - \min_{p\in[|\mathcal{N}_p|]} \frac{\min_{N_p \leq i < \max\{N_{p+1}, M-N\}} V_i}{V_{N_p}}.$$

Fourth, we compute the portfolio turnover to estimate the amount of trading necessary to rebalance each portfolio over time, defined by the average of the ℓ_1-norm of the trades across the I assets:

$$\text{Turnover} = \frac{1}{M-N-1} \sum_{t=1}^{M-N-1} \|\boldsymbol{x}^*_{N+t+1} - \boldsymbol{x}_{N+t^+}\|_1,$$

where $\boldsymbol{x}^*_{N+t+1}$ is the optimal portfolio weight vector at time $N+t+1$ and $\boldsymbol{x}_{N+t^+} \in \mathbb{R}^I$ is the portfolio weight vector right before rebalancing at time $N+t+1$, i.e. for each component $i \in [I]$ of $\boldsymbol{x}_{N+t^+}$, $(\boldsymbol{x}_{N+t^+})_i := \frac{(\boldsymbol{x}_{N+t})_i(1+(\boldsymbol{r}_{N+t})_i)}{\sum_{j\in[I]}(\boldsymbol{x}_{N+t})_j(1+(\boldsymbol{r}_{N+t})_j)} \in \mathbb{R}$.

EC.6. Additional Empirical Results.

EC.6.1. Empirical Results of DeMiguel et al. (2009)'s Datasets from 1997 to 2019

Since the studies in DeMiguel et al. (2009) are up to 2004, it is interesting to continue the study from 1997 to the first quarter of 2019 such that it covers 2008 financial crisis. We still consider the same assets in those six datasets and conduct a similar study. The only difference is that the data used are recent. Table EC.1 presents the empirical results.

Similarly, the RSDR CVaR (HMM) still uniformly outperforms the EW portfolio in all datasets in terms of Sharpe ratio and CEQ. The maximum drawdowns are comparable among all three portfolios. Again, the EW portfolio contains the least turnover but for some datasets, the turnover rates of the EW and RSDR CVaR (HMM) portfolios are comparable. It is noteworthy that from Table EC.1, RS CVaR (HMM) and DR CVaR (Wasserstein) perform well in some datasets but perform badly in others. RSDR CVaR (HMM) merges their merits to perform generally well in all datasets across different metrics.

The DR and RSDR CVaR portfolios in Tables 2 and EC.1 are obtained using the Wassterstein metric with ℓ_1-norm. In Appendix EC.6.2, we present another set of results of RSDR CVaR (HMM) using the Wassterstein distance with ℓ_2-norm. One can observe that the choice of norms does not make a significant difference to the results and our interpretation. Thus, we will stick with ℓ_1-norm.

In summary, our proposed RSDR CVaR (HMM) beats the EW portfolio in terms of Sharpe ratio and CEQ. It also manifests the key feature of regime-switching model when the regime of the market obviously switches, while enjoying the robustness brought from the DRO framework. To better visualize how the wealth changes over time, we report the time series plots of the portfolio wealth with an initial wealth of unit 1 for the three portfolios in Appendix EC.6.3.

The performance results between regime-switching and non-regime-switching case for CVaR models are similar because all assets in each dataset in DeMiguel et al. (2009) are from the same class. Moreover, assets within one class tend to perform more similarly under the state evolution compared with assets between classes. In the next part, we will show that our HMM models would outperform when the given dataset contains assets from different classes.

Strategy	S&P sectors $I=11$	Industry sectors $I=11$	Inter'l portfolios $I=9$	Mkt/ SMB/HML $I=3$	FF 1-factor $I=21$	FF 4-factor $I=24$
1. Sharpe ratio						
EW	0.1374	0.1802	(0.0873)	**0.0730**	0.1485	0.1462
RS CVaR (HMM)	**0.1637**	**0.2642**	**(0.0845)**	(0.0644)	**0.1689**	(0.0409)
DR CVaR (Wasserstein)	0.1413	0.1959	(0.0873)	**0.0730**	0.1544	0.1555
RSDR CVaR (HMM)	0.1464	0.1989	(0.0873)	**0.0730**	0.1557	**0.1630**
2. CEQ						
EW	0.0047	0.0068	(0.0056)	**0.0014**	0.0064	**0.0057**
RS CVaR (HMM)	0.0044	**0.0085**	**(0.0043)**	(0.0015)	0.0064	(0.0009)
DR CVaR (Wasserstein)	0.0048	0.0069	(0.0056)	**0.0014**	**0.0067**	0.0055
RSDR CVaR (HMM)	**0.0049**	0.0072	(0.0056)	**0.0014**	**0.0067**	**0.0057**
3. Maximum drawdown						
EW	0.4986	0.4848	0.0873	0.2862	0.5380	0.4860
RS CVaR (HMM)	**0.4034**	**0.3455**	**0.0846**	**0.2091**	0.5249	**0.1951**
DR CVaR (Wasserstein)	0.5010	0.4438	0.0873	0.2868	**0.5214**	0.4345
RSDR CVaR (HMM)	0.4929	0.4488	0.0873	0.2859	0.5216	0.4268
4. Turnover						
EW	**0.0291**	**0.0261**	**0.0253**	**0.0247**	**0.0207**	**0.0246**
RS CVaR (HMM)	0.2511	0.3650	0.3446	0.0821	0.3259	0.1902
DR CVaR (Wasserstein)	0.0783	0.0564	**0.0253**	**0.0247**	0.0814	0.1781
RSDR CVaR (HMM)	0.1055	0.0894	**0.0253**	**0.0247**	0.1794	0.1951

Table EC.1 Empirical Results of DeMiguel et al. (2009)'s datasets from 1997 to 2019 Q1.

EC.6.2. Empirical Results with ℓ_2-norm in Wasserstein Distance

For the cost function $\rho(\mathbf{p},\mathbf{q}) = \|\mathbf{p}-\mathbf{q}\|_m$ in (2), we apply ℓ_2-norm (i.e. $m=2$) to allow generalizations of our regime-switching DRO model. The results of ℓ_2-norm for two time periods are shown in Tables EC.2 and EC.3, compared with Tables 2 and EC.1, respectively. Generally, when we consider the same hyperparameter selection for the radius of the ambiguity set θ, we find that the performance of ℓ_2-norm is similar to that of ℓ_1-norm but requires much more computation in each oracle.

Metric	S&P sectors $I=11$	Industry sectors $I=11$	Inter'l portfolios $I=9$	Mkt/ SMB/HML $I=3$	FF 1-factor $I=21$	FF 4-factor $I=24$
Sharpe Ratio	0.1838	0.1375	0.1270	0.2259	0.1664	0.1869
CEQ	0.0067	0.0050	0.0046	0.0039	0.0074	0.0072
Maximum drawdown	0.3195	0.4459	0.3907	0.1711	0.4349	0.2729
Turnover	0.0424	0.0287	0.0380	0.0265	0.0268	0.0346

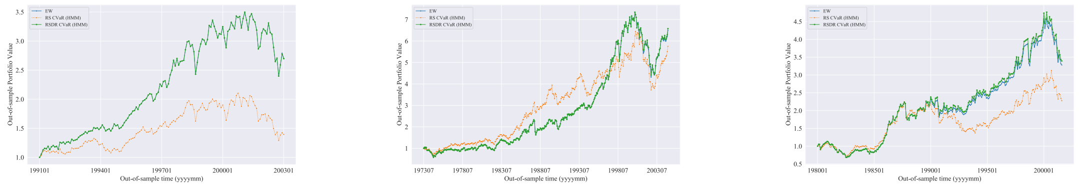
Table EC.2 Empirical Results of DeMiguel et al. (2009)'s datasets during the same time with ℓ_2-norm.

Metric	S&P sectors $I=11$	Industry sectors $I=11$	Inter'l portfolios $I=9$	Mkt/ SMB/HML $I=3$	FF 1-factor $I=21$	FF 4-factor $I=24$
Sharpe Ratio	0.1449	0.1892	-0.0870	0.0690	0.1515	0.1539
CEQ	0.0049	0.0069	-0.0054	0.0013	0.0065	0.0049
Maximum Drawdown	0.4891	0.4556	0.6193	0.2574	0.5282	0.3661
Turnover	0.0523	0.0470	0.0402	0.0256	0.0603	0.0950

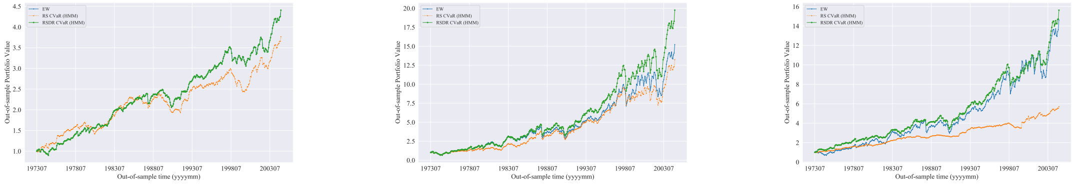
Table EC.3 Empirical Results of DeMiguel et al. (2009)'s datasets from 1997 to 2019 Q1 with ℓ_2-norm.

EC.6.3. Time Series Plots of the Portfolio Wealth for the Empirical Studies

Figures below show how an initial wealth of unit one varies over time with different portfolios and in different time periods. Figure EC.1 corresponds to the empirical results in Section 5.2 and Figure EC.2 corresponds to the empirical results in Appendix EC.6.1. Note that in some cases, RSDR CVaR (HMM) models is the same as EW portfolio across time and the blue dotted line coincides with the green solid line. We find that our proposed RSDR CVaR (HMM) models can achieve the most portfolio wealth across time periods under most cases.

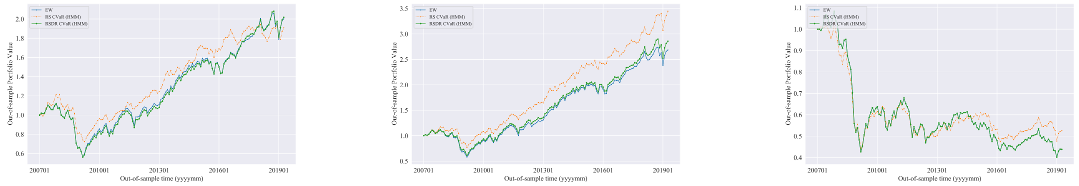


(a) S&P sectors ($I = 11$) (b) Industry sectors ($I = 11$) (c) Inter'l portfolios ($I = 9$)

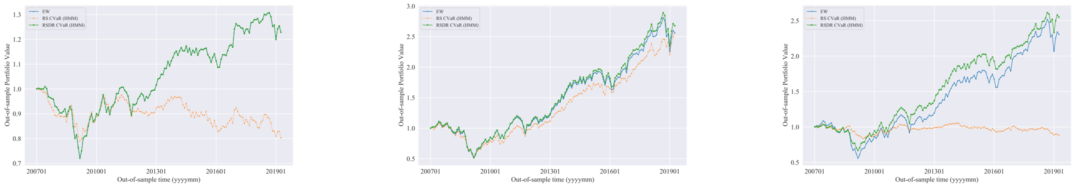


(d) Mkt/SMB/HM ($I = 3$) (e) FF 1-factor ($I = 21$) (f) FF 1-factor ($I = 24$)

Figure EC.1 Portfolio values of DeMiguel et al. (2009)'s datasets during the same time



(a) S&P sectors ($I = 11$) (b) Industry sectors ($I = 11$) (c) Inter'l portfolios ($I = 9$)



(d) Mkt/SMB/HM ($I = 3$) (e) FF 1-factor ($I = 21$) (f) FF 1-factor ($I = 24$)

Figure EC.2 Portfolio values of DeMiguel et al. (2009)'s datasets from 1997 to 2019

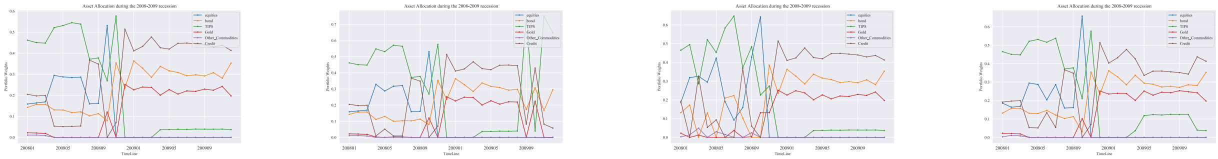
Figures below show how an initial wealth of unit one varies over time corresponding to the empirical results in Section 5.3. Figure EC.3 corresponds to the empirical results of the Weathers and the Bull & Bear Approach.

(a) CVaR (Weathers) (b) Mean-CVaR (Weathers) (c) CVaR (Bull & Bear) (d) Mean-CVaR (Bull & Bear)

Figure EC.3 Comparison of different Weathers and Bull & Bear models.

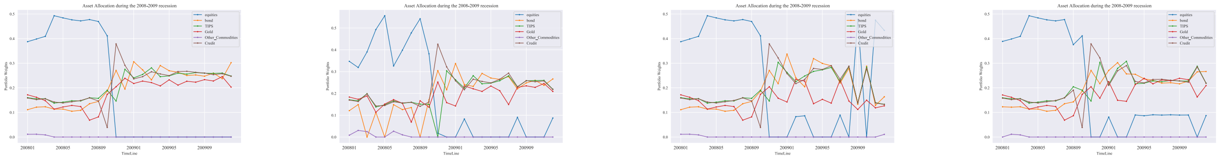
EC.6.4. Asset Allocation for Different Models during the 2008-2009 Great Recession

To investigate the effects of parameter sensitivity and regime determination across RSDR CVaR models, we compare different ambiguity set sizes θ for DR-CVaR (Wasserstein), RSDR-CVaR (HMM), RSDR-CVaR (Bull&Bear), and RSDR-CVaR (Weathers) during the 2008-2009 recession in Figures EC.4 ($\gamma = 0$), EC.5 ($\gamma = 0.005$), EC.6 ($\gamma = 0.01$),and EC.7 ($\gamma = 0.025$). In particular, when $\gamma = 0$, the distributionally robust models reduce to the non-robust counterpart.



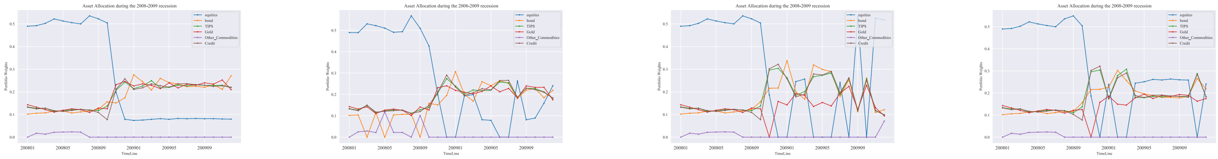
(a) CVaR (b) RS CVaR (HMM) (c) RS CVaR (Weathers) (d) RS CVaR (Bull&Bear)

Figure EC.4 Asset Allocation for different models during the 2008-2009 great recession when $\gamma = 0$



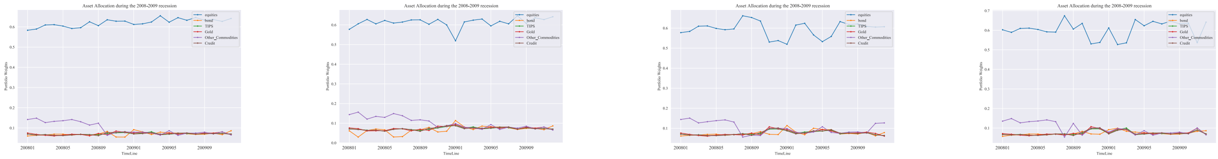
(a) DR (b) RSDR (HMM) (c) RSDR (Weathers) (d) RSDR (Bull&Bear)

Figure EC.5 Asset Allocation for different models during the 2008-2009 great recession when $\gamma = 0.005$



(a) DR (b) RSDR (HMM) (c) RSDR (Weathers) (d) RSDR (Bull&Bear)

Figure EC.6 Asset Allocation for different models during the 2008-2009 great recession when $\gamma = 0.01$



(a) DR (b) RSDR (HMM) (c) RSDR (Weathers) (d) RSDR (Bull&Bear)

Figure EC.7 Asset Allocation for different models during the 2008-2009 great recession when $\gamma = 0.05$

One can see that when we increase the tuning parameter m to 0.025, the asset allocation would have little variation across time periods (and thus increasing m is somehow practically meaningless). For each fixed m (or equivalently θ), we can observe that the RSDR-CVaR models are generally more volatile than the DR-CVaR model over time, which indicates that regime-switching model considers different mixtures of underlying distributions for asset returns over time and thus can capture the time-varying uncertainties. The comparison between DR and RSDR models is subtle, which can be ascribed to the small window size we chose. When m is not large (i.e., the proposed RSDR models do not converge to EW portfolio), our RSDR models can identify regime-switching behavior based on the negative returns of equities and excess positive returns of bond.