Algorithm 1 Systematic estimation for ϵ and ρ

Require: independent empirical **score** distributions $\hat{\mathbb{P}}_n^1, \hat{\mathbb{P}}_n^2$ from original distribution $\mathbb{P}$, empirical **score** distribution $\hat{\mathbb{Q}}_m$ from shifted distribution $\mathbb{Q}$ $(m < n)$, a grid of $\epsilon_{i\in[k]}$, target coverage $1-\alpha$.

Ensure: pair of ϵ_i, ρ_i with tightest prediction set and asymptotic valid coverage $1-\alpha$.

1: **for** $i = 1, \cdots, k$, **do**

2: Compute OT distance ρ_i between $\mathbb{P}_n^1$ and $\hat{\mathbb{Q}}_m$ with transport cost $\mathbb{1}\{|x-y| \geq \epsilon_i\}$

3: Compute empirical worst-case quantile $q_i = \text{Quant}^{\text{WC}}_{\epsilon_i,\rho_i}(1-\beta; \hat{\mathbb{P}}_n^2) := \text{Quant}(1-\beta+\rho_i; \hat{\mathbb{P}}_n^2)+\epsilon_i$, where $\beta_i = \alpha+(\alpha-\rho_i-2)/n$

4: **end for**

5: **return** (ϵ_i, ρ_i) with minimal q_i ▷ Bigger worst-case quantile q_i implies bigger size of prediction sets

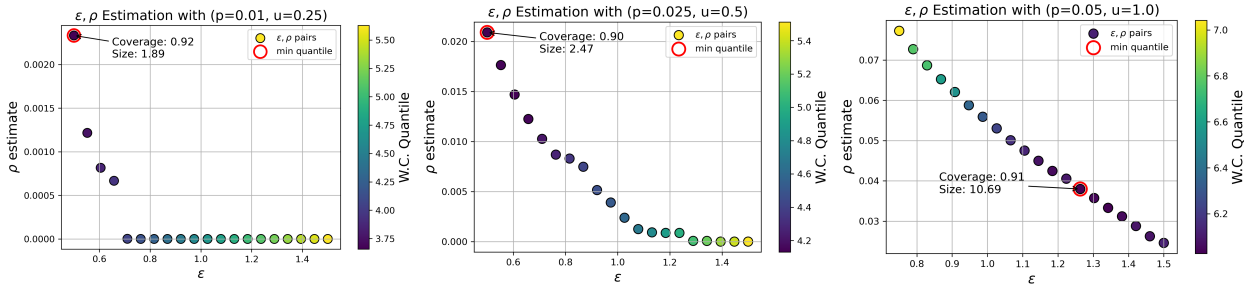

Figure 3: **ImageNet ε, ρ estimation plot**. Here, (p, u) corresponds to the true corruption ratios and noise levels with respective estimates (ρ, ε). The plot contains a 20-point ε-ρ grid, with $\varepsilon \in (0.5, 1.5)$ and ρ estimated using optimal transport. The empirical worst-case quantile is visualized via colorbars. The optimal estimated (ε, ρ) pair is highlighted in red, with empirical coverage and prediction set size annotated.

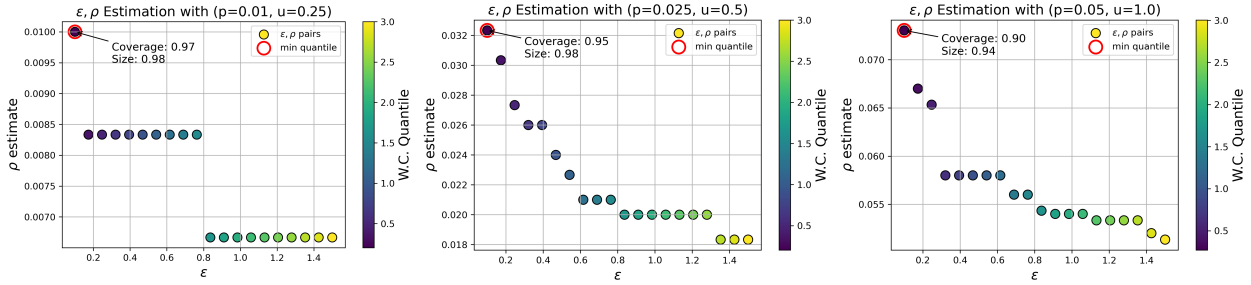

Figure 4: **MNIST ε, ρ estimation plot**. Here, (p, u) corresponds to the true corruption ratios and noise levels with respective estimates (ρ, ε) The plot contains a 20-point ε-ρ grid, with $\varepsilon \in (0.1, 1.5)$ and ρ estimated using optimal transport. The empirical worst-case quantile is visualized via colorbars. The optimal estimated (ε, ρ) pair is highlighted in red, with empirical coverage and prediction set size annotated.

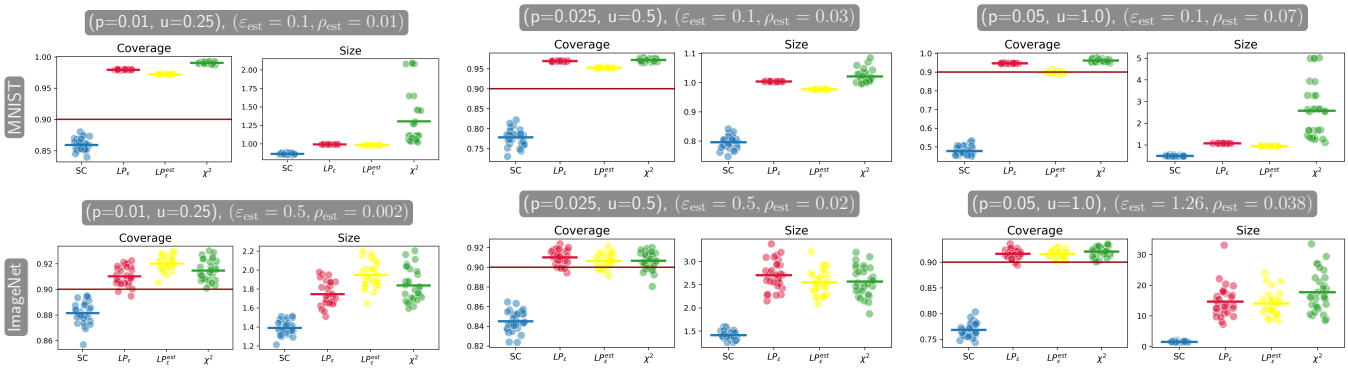

Figure 5: **Data-space shift validity and efficiency.** Here, (p, u) corresponds to the true corruption ratio and noise level with respective estimates $(\rho_{\text{est}}, \varepsilon_{\text{est}})$. Desired $1-\alpha$ coverage (long dark red line); empirical coverage and prediction set size for each split (scattered points); and mean coverage and prediction set size across 30 calibration-test splits (short colored horizontal lines) are plotted. We include one extra setting (LP_ε^{est}, yellow scatter points) where we estimate ε, ρ using Algorithm 1.