**Scenario 2 – No Filters (Outliers Removed)**

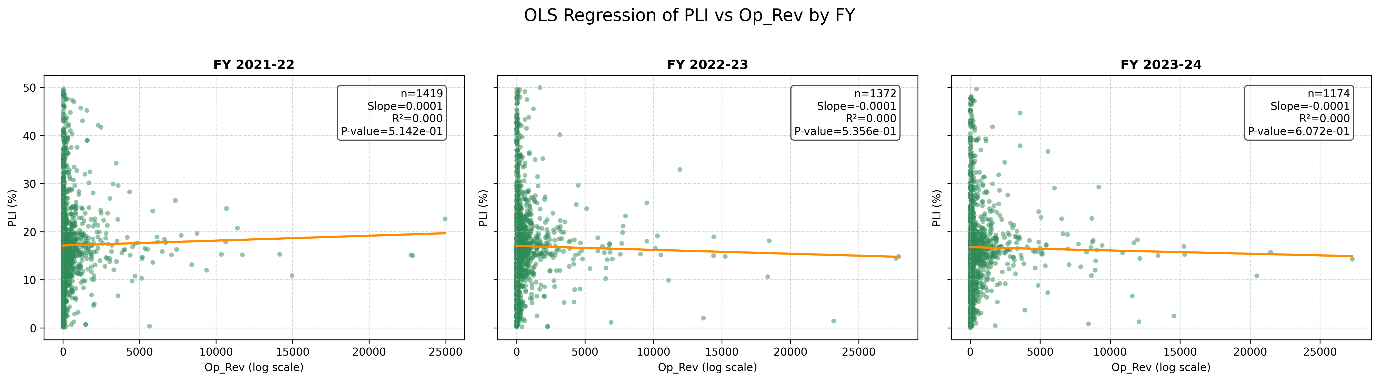
**Scenario description**

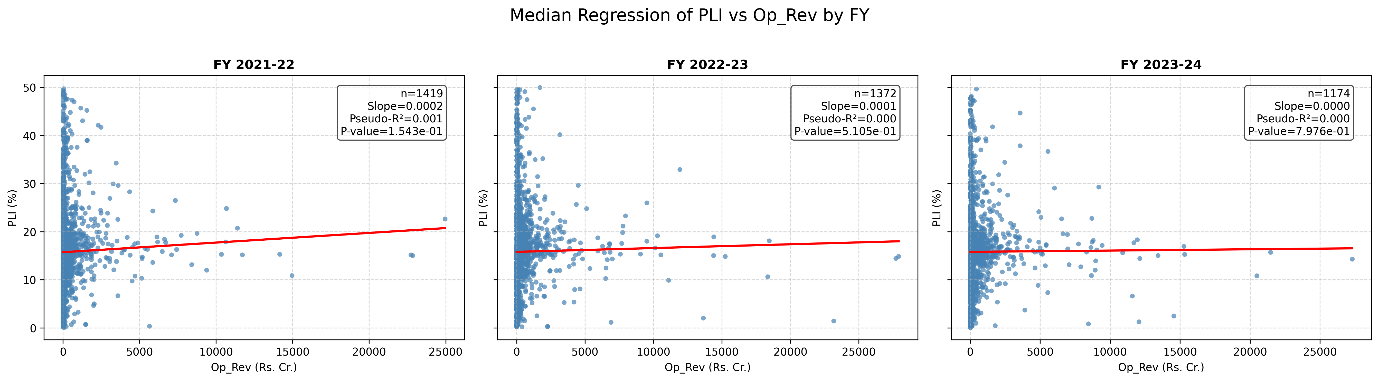
This scenario mirrors the baseline (Scenario 1) but additionally excludes statistical outliers in revenue and profitability. The applied filters are:

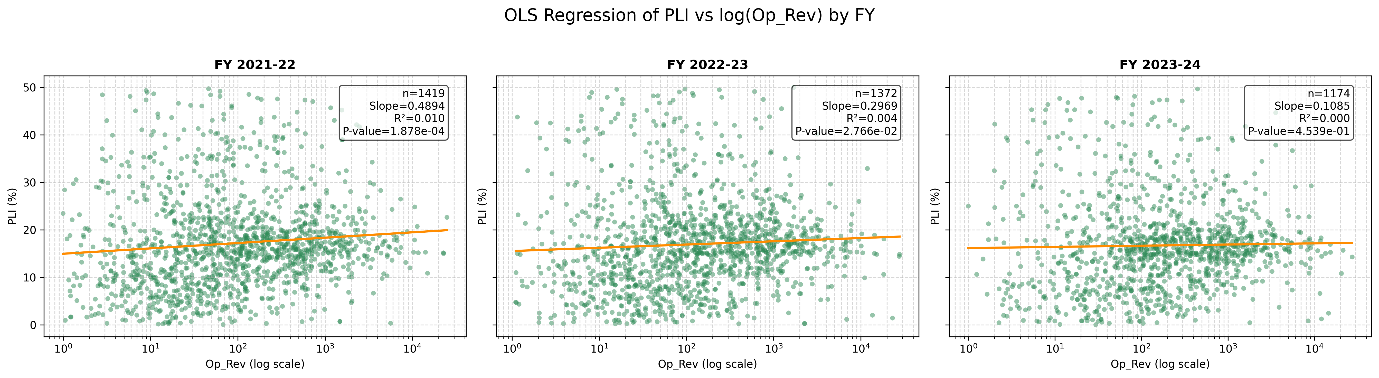
* **Net worth > 0**
* **Persistent loss = 0**
* **Employee cost ≥ 25% of OR**
* **Service income ≥ 75% of OR**
* **Operating revenue between 1 and 30,000 Cr**
* **PLI between 0 and 50**

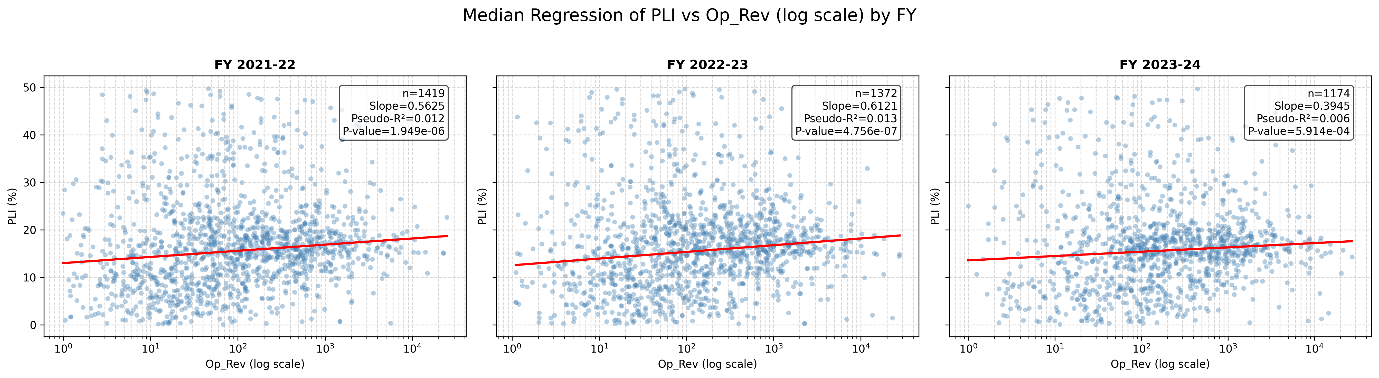
Final sample sizes: **FY 2021-22: 1,419**, **FY 2022-23: 1,372**, **FY 2023-24: 1,174** firm-year observations.

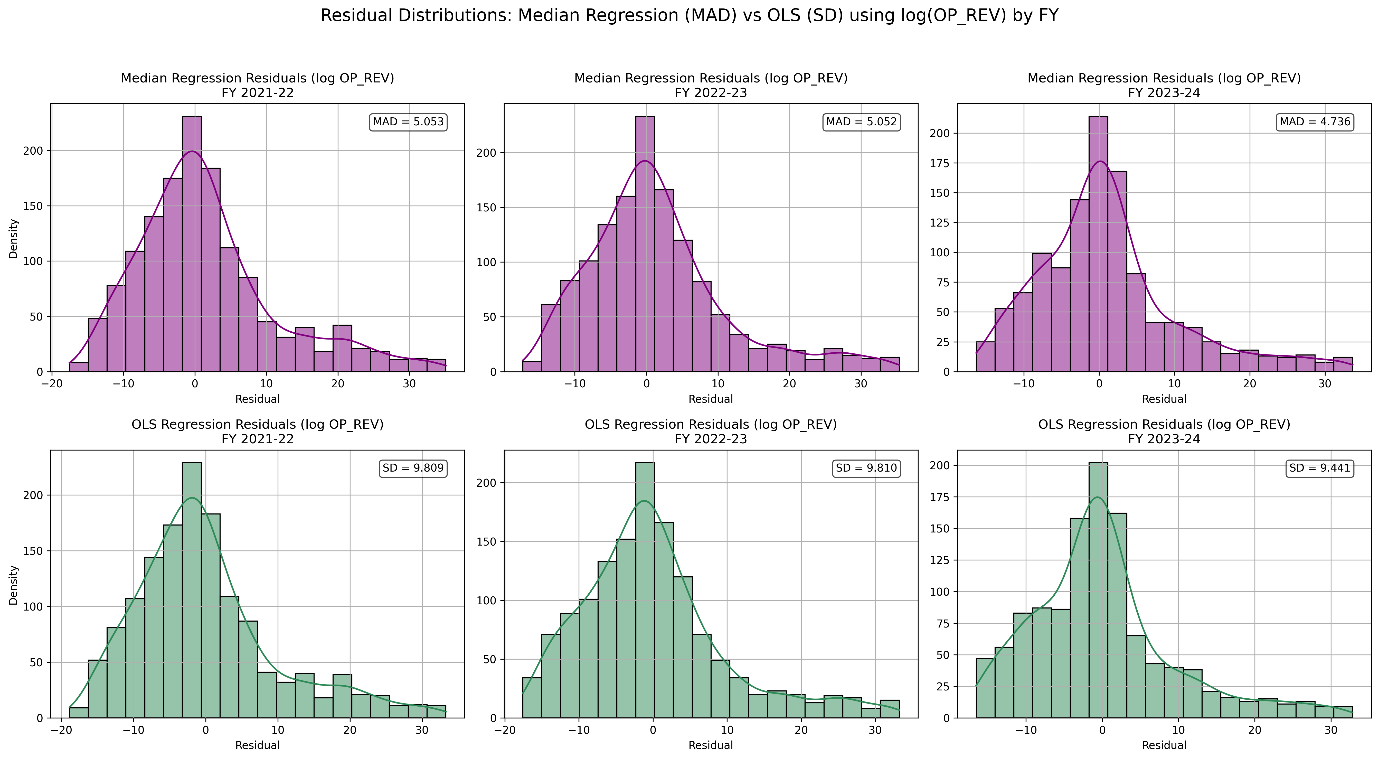
**Results:**











**Interpretation:**

**Direct OR vs PLI (linear form)**

**OLS regression**

* R² values remain close to zero in all years.
* Slopes are statistically insignificant (p-values 0.51–0.61).
* **Inference:** No meaningful relationship detected between OR and profitability.

**Median regression**

* Slopes are very small (0.00003–0.00020) and **not statistically significant** (p-values 0.15–0.80).
* Pseudo-R² values are negligible (<0.001).
* **Inference:** After removing outliers, even the weak statistical significance observed earlier disappears. Direct OR does not explain profitability.

**Log(OR) vs PLI (functional transformation)**

**OLS regression**

* Slopes are positive in all years, with significance in FY 2021-22 (p < 0.001) and FY 2022-23 (p ≈ 0.028), but not in FY 2023-24 (p ≈ 0.45).
* R² values remain modest (0.003–0.010).
* **Inference:** OLS detects some association in earlier years, but the effect is unstable and weak.

**Median regression**

* Slopes are positive and **highly significant across all years** (p < 0.001).
* Pseudo-R² improves (0.006–0.013), clearly higher than Scenario 1.
* Residual MAD (≈4.7–5.0) is much lower than OLS residual SD (≈9.4–9.8), confirming robustness.
* **Inference:** Median regression with log(OR) remains the strongest model, with explanatory power improving once outliers are removed. Larger firms tend to exhibit higher profitability on a median basis.

**Overall takeaway for Scenario 2**

* Outlier removal eliminates spurious significance in the direct OR–PLI relationship.
* **Median regression with log(OR)** continues to be the most reliable specification, now showing **clearer and stronger associations** than in Scenario 1.

**Prediction Table:**

Based on the analysis, the median regression with log(OR) emerges as the most significant and robust model. Accordingly, the following table presents predicted PLI values generated from this model for different levels of Operating Revenue (Op\_Rev).

