**Scenario 1 – No Filters**

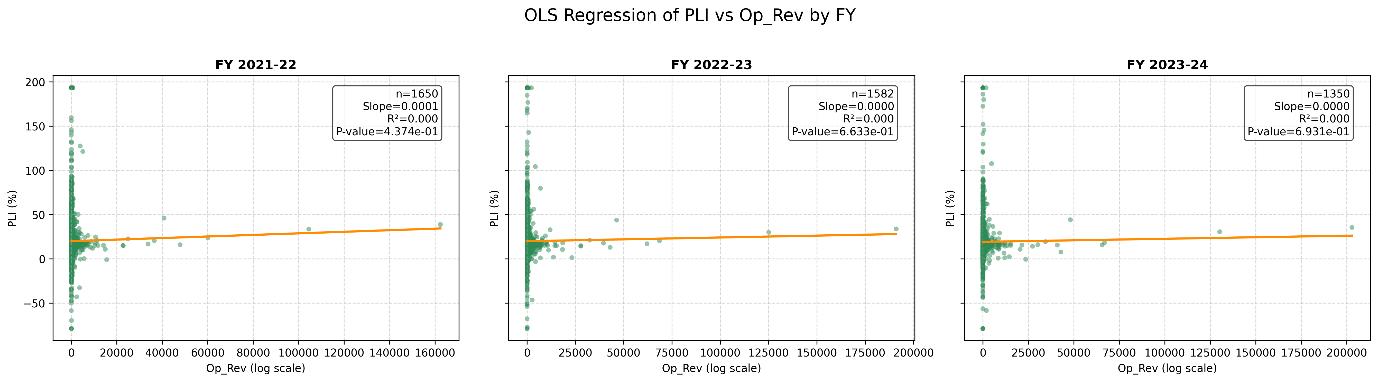
**Scenario description**

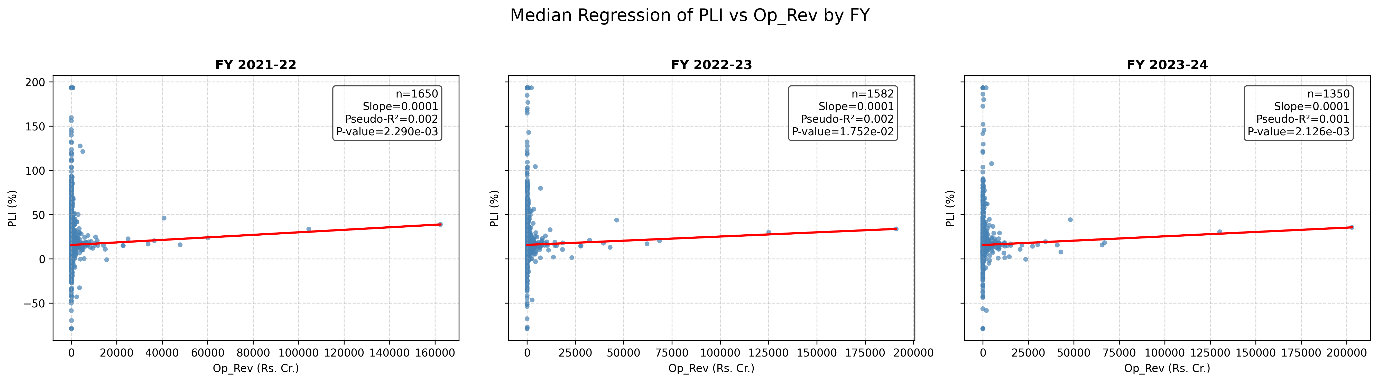
This baseline scenario includes the **full set of IT/ITeS companies** without excluding any firms for export orientation or related party transactions (RPTs). Only the following basic eligibility filters were applied:

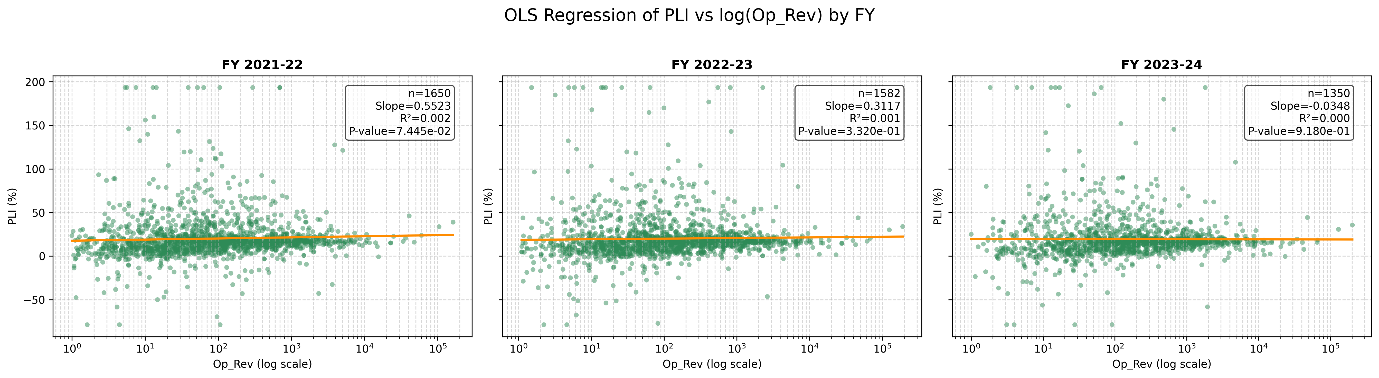
* **Net worth > 0**
* **Persistent loss = 0** (i.e., excluded firms with negative operating profit across all years)
* **Employee cost ≥ 25% of OR** (ensuring service intensity)
* **Service income ≥ 75% of OR** (ensuring predominantly IT/ITeS activity)

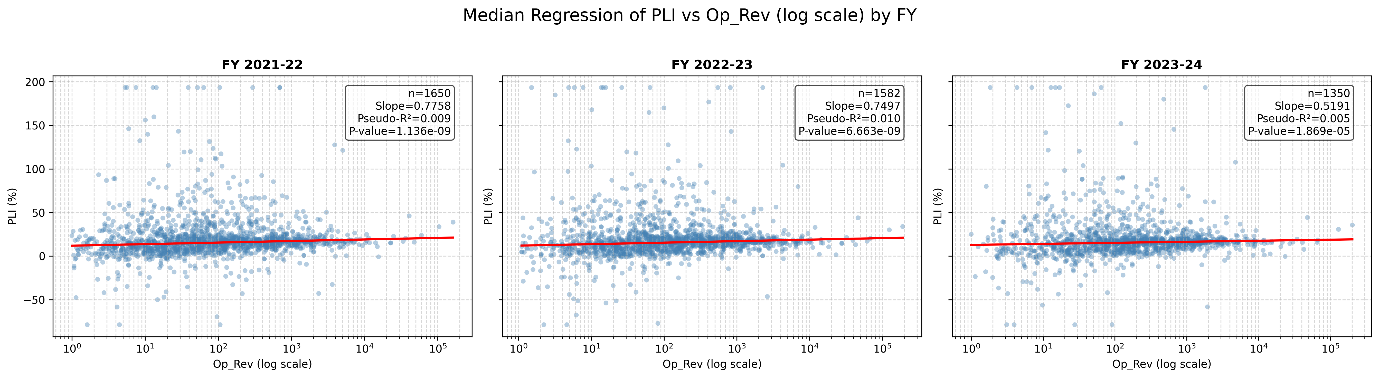
Final sample sizes: **FY 2021-22: 1,650**, **FY 2022-23: 1,582**, **FY 2023-24: 1,350** firm-year observations.

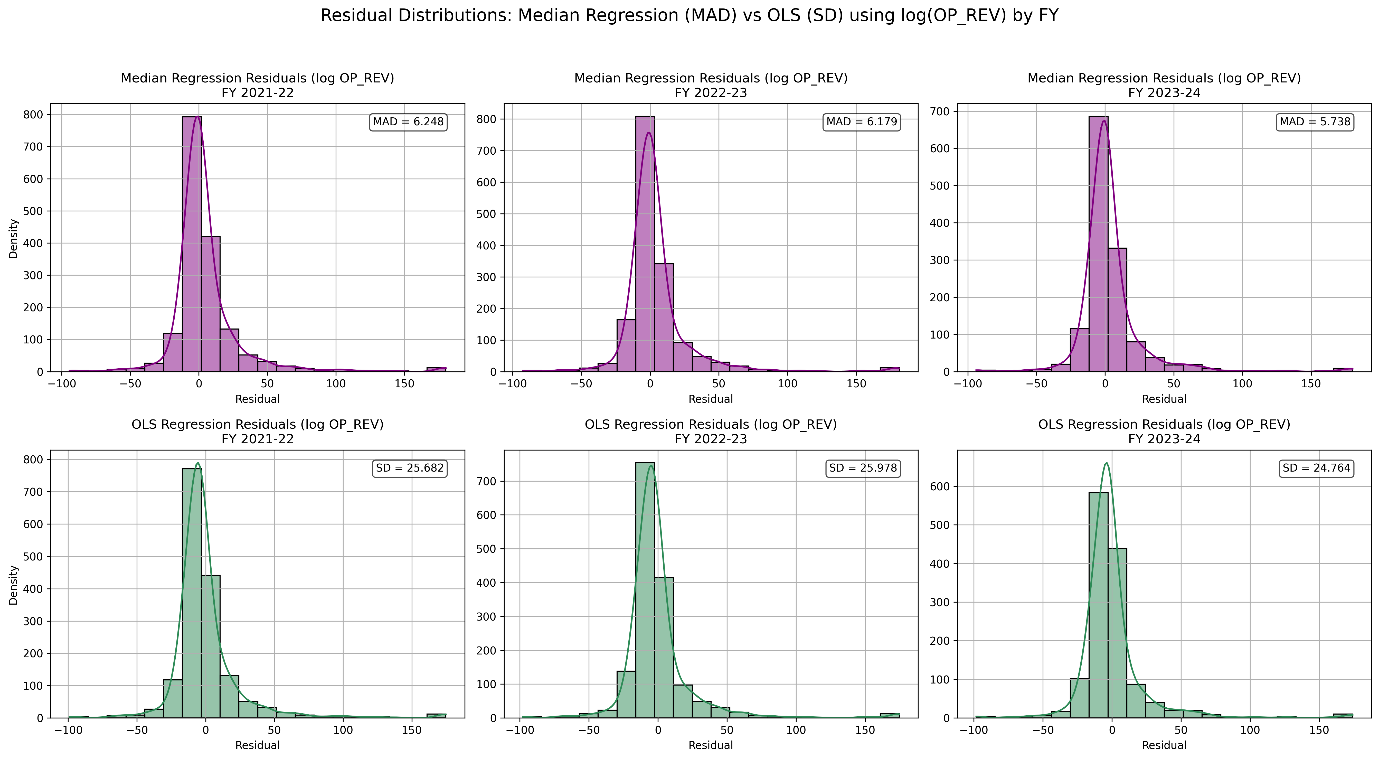
**Results:**











**Interpretation:**

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

**OLS regression**

* Results show **extremely weak fit** (R² < 0.001 in all years).
* Slopes are statistically insignificant (p-values 0.44–0.69).
* **Inference:** OLS finds no meaningful linear association between operating revenue and profitability.

**Median regression**

* Slopes are very small (≈0.0001) but statistically significant (p < 0.02).
* Pseudo-R² is negligible (<0.003).
* **Inference:** Although statistically detectable, the relationship is too weak to have substantive explanatory value.

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

**OLS regression**

* Slopes are inconsistent (positive in two years, negative in one).
* Statistical significance is absent (p-values 0.07–0.92).
* R² values are near zero.
* **Inference:** Even with log transformation, OLS fails to capture any robust link between firm size and profitability.

**Median regression**

* Slopes are positive and highly significant in all years (p < 1e-5).
* Pseudo-R² improves to the 0.005–0.01 range.
* Residual dispersion (MAD ≈ 5.7–6.2) is far lower than OLS residual SD (≈24–26), confirming robustness.
* **Inference:** Median regression on log(OR) clearly provides the most stable and interpretable model — larger firms (higher log(OR)) tend to achieve higher profitability on a median basis.

**Overall takeaway for Scenario 1**

* **Median regression with log(OR)** outperforms all other specifications.
* The relationship, while modest in explanatory power, is **statistically significant, robust to outliers, and economically interpretable**.

**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).

