

## 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:

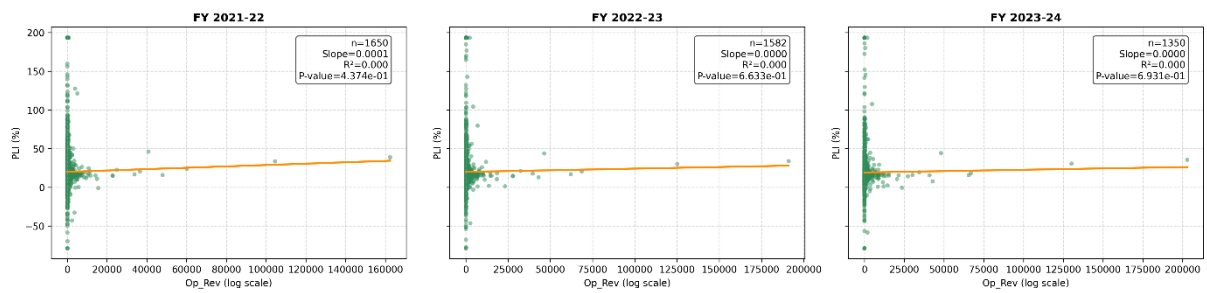
- **Net worth > 0**
- **Persistent loss = 0** (i.e., excluded firms with negative operating profit across all years)
- **Employee cost ≥ 25% of OR** (ensuring service intensity)
- **Service income ≥ 75% of OR** (ensuring predominantly IT/ITeS activity)

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

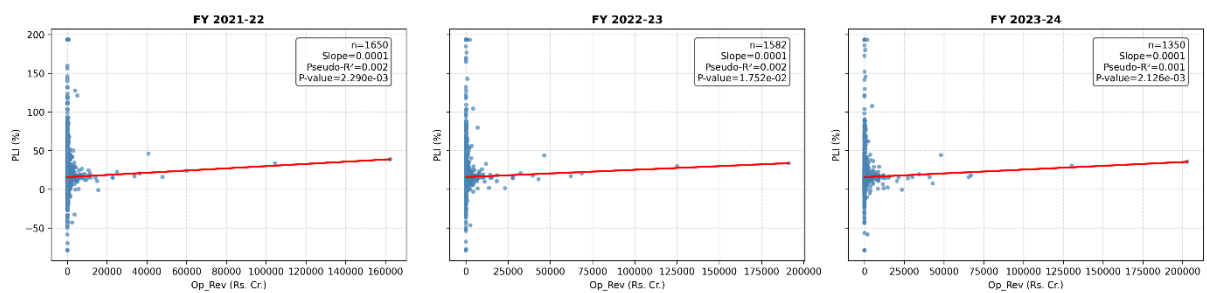
---

## Results:

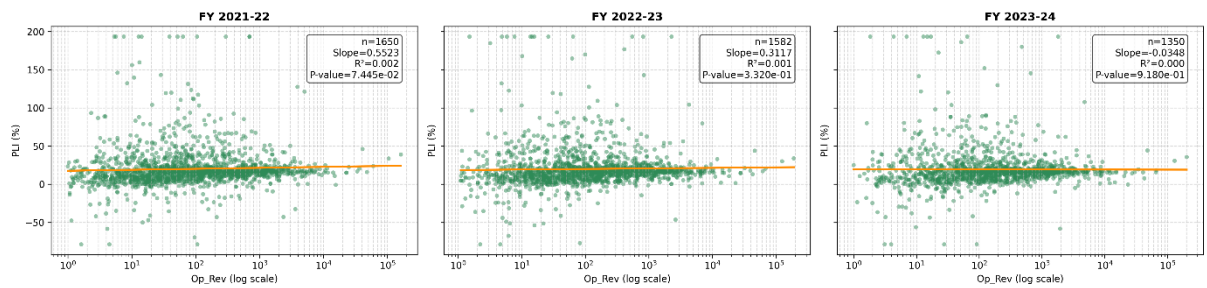
OLS Regression of PLI vs Op\_Rev by FY



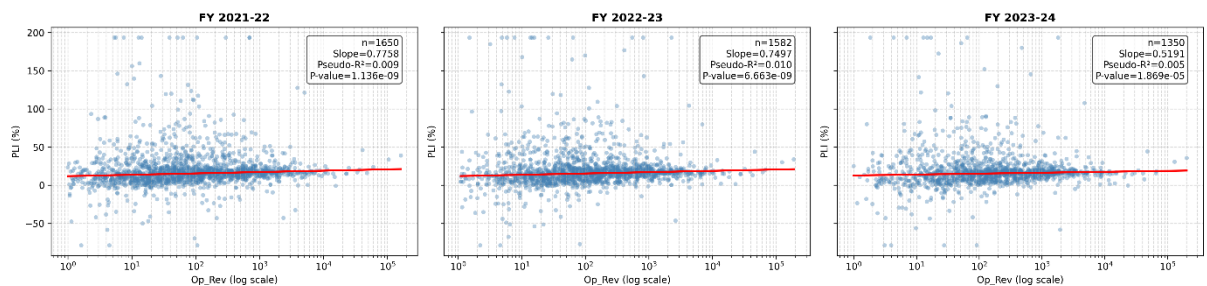
Median Regression of PLI vs Op\_Rev by FY



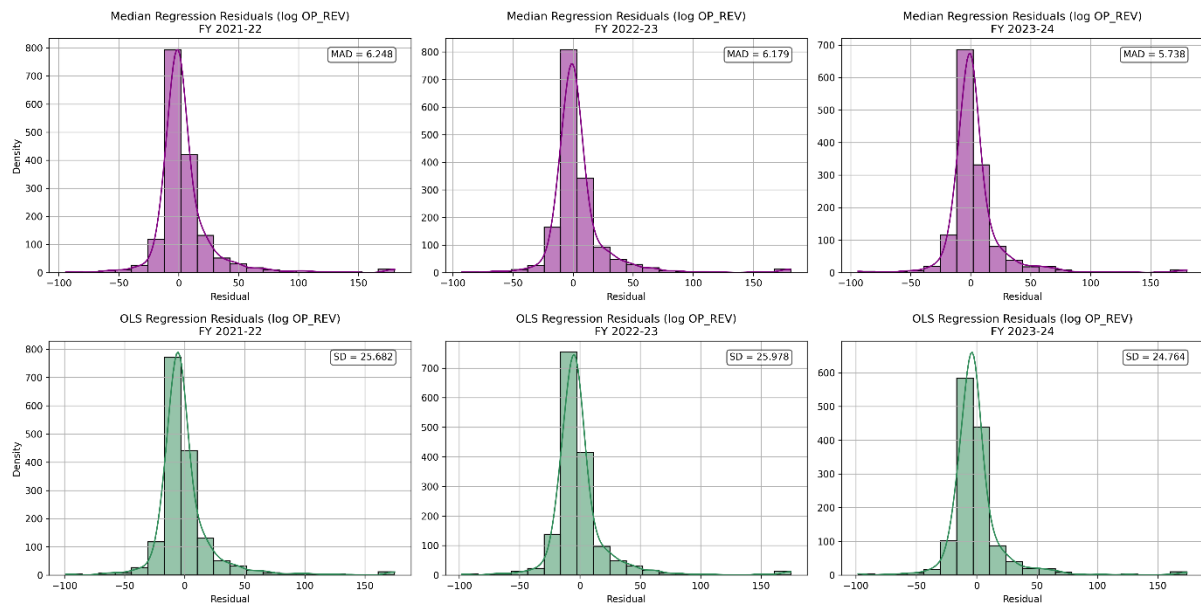
OLS Regression of PLI vs log(Op\_Rev) by FY



Median Regression of PLI vs Op\_Rev (log scale) by FY



### Residual Distributions: Median Regression (MAD) vs OLS (SD) using log(OP\_REV) by FY



### Interpretation:

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

##### OLS regression

- Results show **extremely weak fit** ( $R^2 < 0.001$  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 ( $\approx 0.0001$ ) but statistically significant ( $p < 0.02$ ).
- Pseudo- $R^2$  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).
- $R^2$  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- $R^2$  improves to the 0.005–0.01 range.
  - Residual dispersion ( $MAD \approx 5.7\text{--}6.2$ ) is far lower than OLS residual SD ( $\approx 24\text{--}26$ ), 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).

FY	2021-22	2022-23	2023-24	Average
OP_REV				
1	15.77	15.87	15.71	15.78
100	15.79	15.88	15.72	15.80
300	15.81	15.90	15.74	15.82
500	15.84	15.92	15.76	15.84
1000	15.91	15.96	15.81	15.89
1500	15.98	16.01	15.86	15.95
2000	16.06	16.06	15.91	16.01
3000	16.20	16.15	16.01	16.12
5000	16.48	16.34	16.20	16.34
10000	17.20	16.81	16.69	16.90
20000	18.62	17.75	17.68	18.02

## **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:

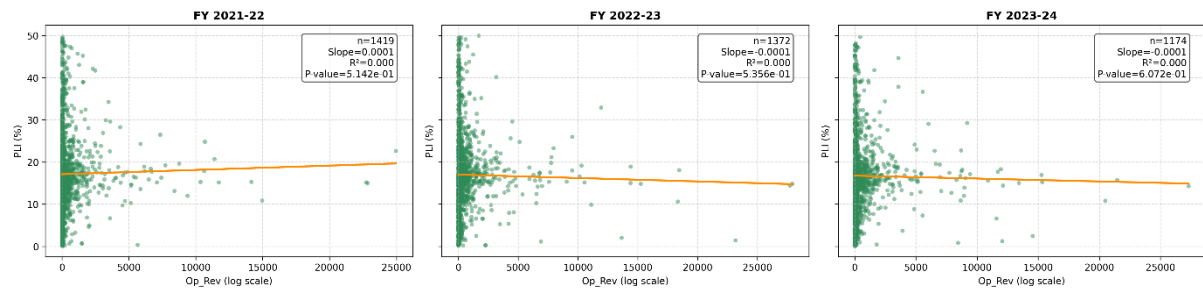
- **Net worth > 0**
- **Persistent loss = 0**
- **Employee cost  $\geq$  25% of OR**
- **Service income  $\geq$  75% of OR**
- **Operating revenue between 1 and 30,000 Cr**
- **PLI between 0 and 50**

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

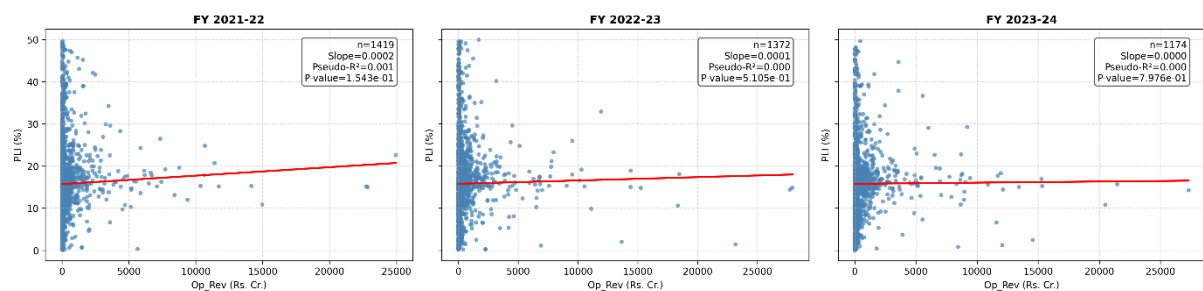
---

## Results:

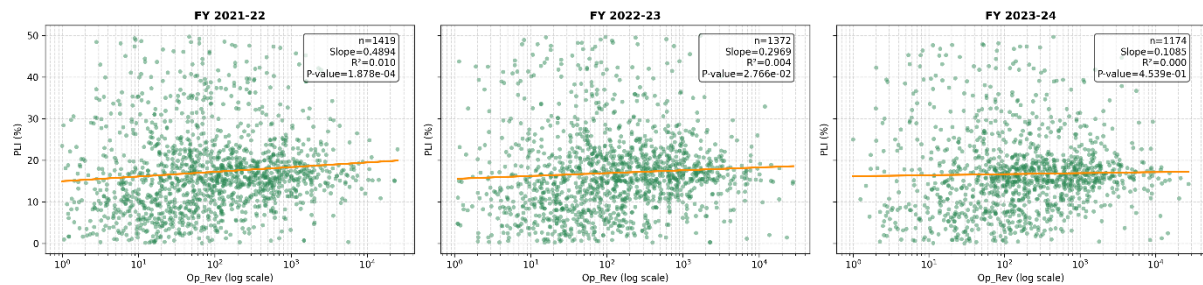
OLS Regression of PLI vs Op\_Rev by FY



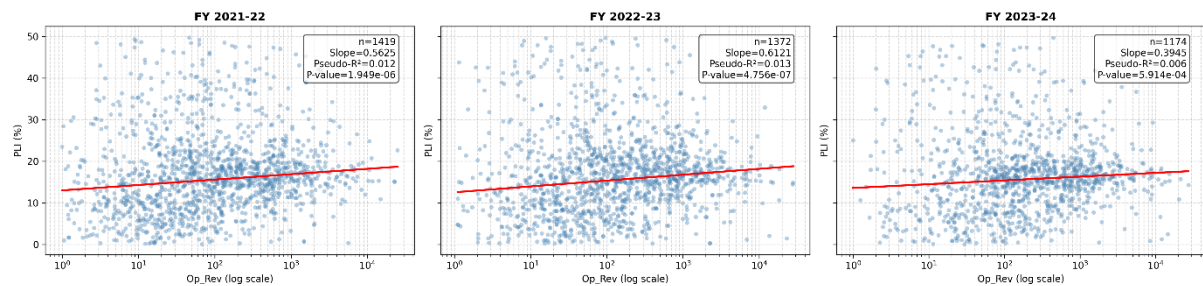
Median Regression of PLI vs Op\_Rev by FY



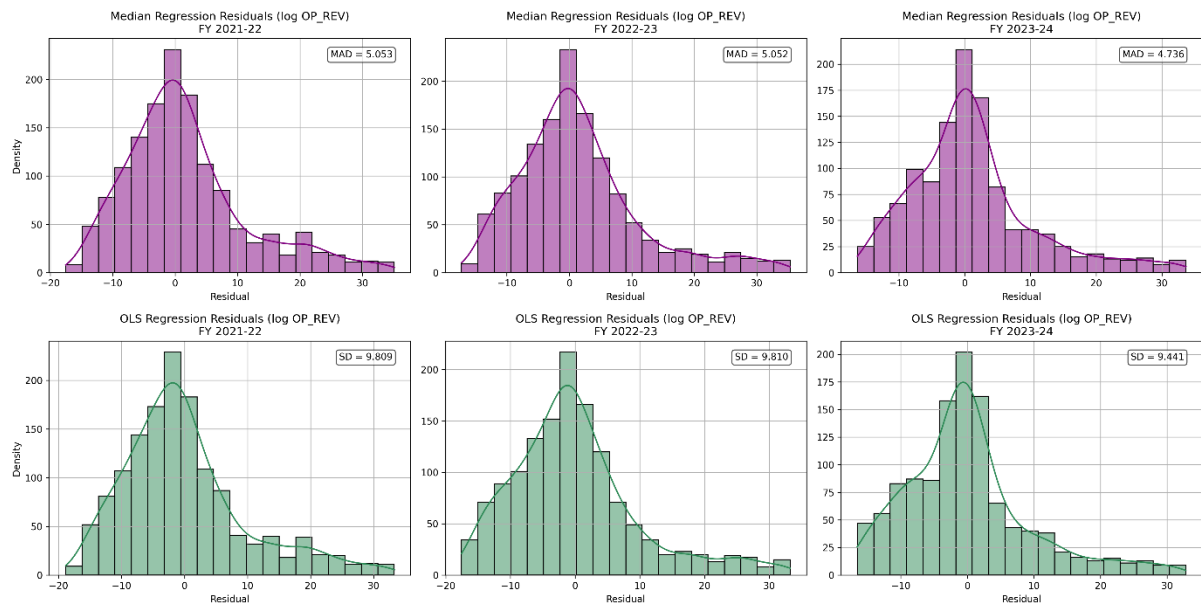
OLS Regression of PLI vs log(Op\_Rev) by FY



Median Regression of PLI vs Op\_Rev (log scale) by FY



### Residual Distributions: Median Regression (MAD) vs OLS (SD) using log(OP\_REV) by FY



### Interpretation:

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

##### OLS regression

- $R^2$  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- $R^2$  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 ( $p \approx 0.028$ ), but not in FY 2023-24 ( $p \approx 0.45$ ).
- $R^2$  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- $R^2$  improves (0.006–0.013), clearly higher than Scenario 1.
- Residual MAD ( $\approx 4.7$ – $5.0$ ) is much lower than OLS residual SD ( $\approx 9.4$ – $9.8$ ), confirming robustness.
- **Inference:** Median regression with  $\log(\text{OR})$  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  $\log(\text{OR})$**  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  $\log(\text{OR})$  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).

FY	2021-22	2022-23	2023-24	Average
OP_REV				
1	13.00	12.54	13.58	13.04
100	15.59	15.36	15.40	15.45
300	16.21	16.04	15.83	16.03
500	16.50	16.35	16.03	16.29
1000	16.89	16.77	16.31	16.66
1500	17.12	17.02	16.47	16.87
2000	17.28	17.20	16.58	17.02
3000	17.51	17.45	16.74	17.23
5000	17.79	17.76	16.94	17.50
10000	18.18	18.18	17.22	17.86
20000	18.57	18.61	17.49	18.22



### Scenario 3 – Export Oriented

#### Scenario description

This scenario applies the same filters as Scenario 2 (removal of outliers) but adds an **export intensity filter**, restricting the sample to export-oriented firms only. No restriction was applied on related party transactions (RPTs). The applied filters are:

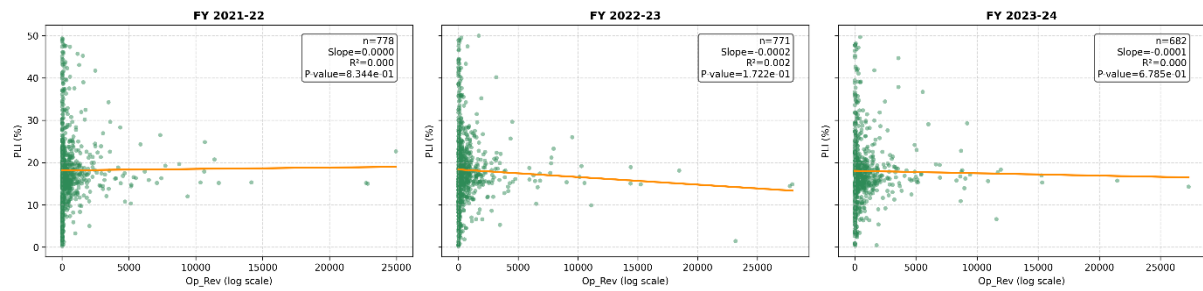
- **Net worth > 0**
- **Persistent loss = 0**
- **Employee cost  $\geq$  25% of OR**
- **Service income  $\geq$  75% of OR**
- **Operating revenue between 1 and 30,000**
- **PLI between 0 and 50**
- **Export income  $\geq$  75% of OR**

Final sample sizes: **FY 2021-22: 778, FY 2022-23: 771, FY 2023-24: 682** firm-year observations.

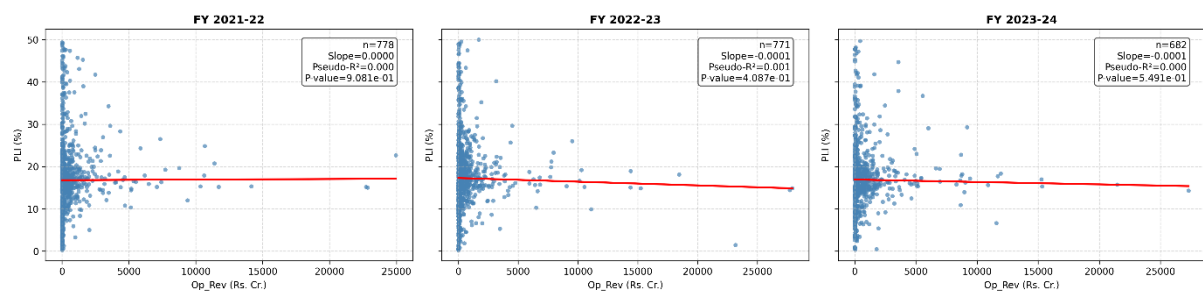
---

## Results:

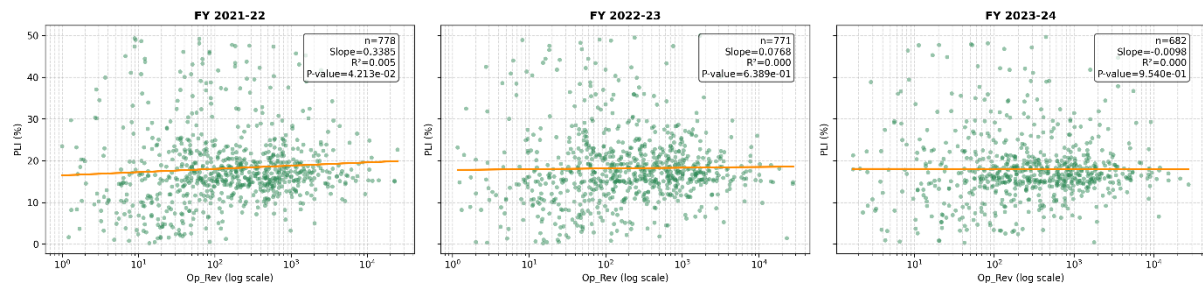
OLS Regression of PLI vs Op\_Rev by FY



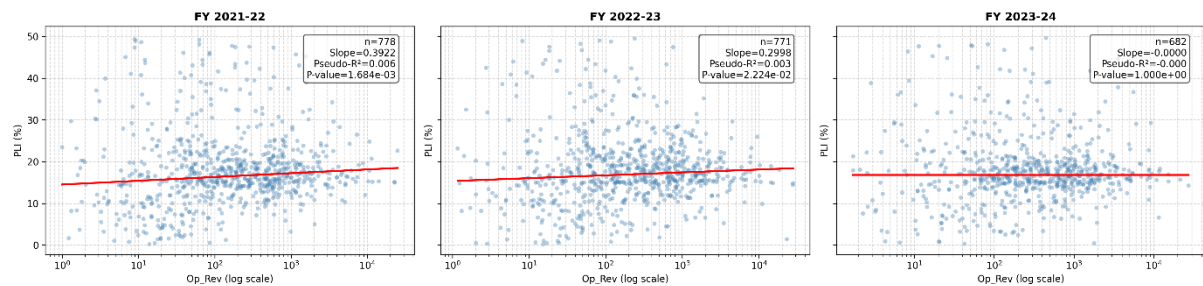
Median Regression of PLI vs Op\_Rev by FY



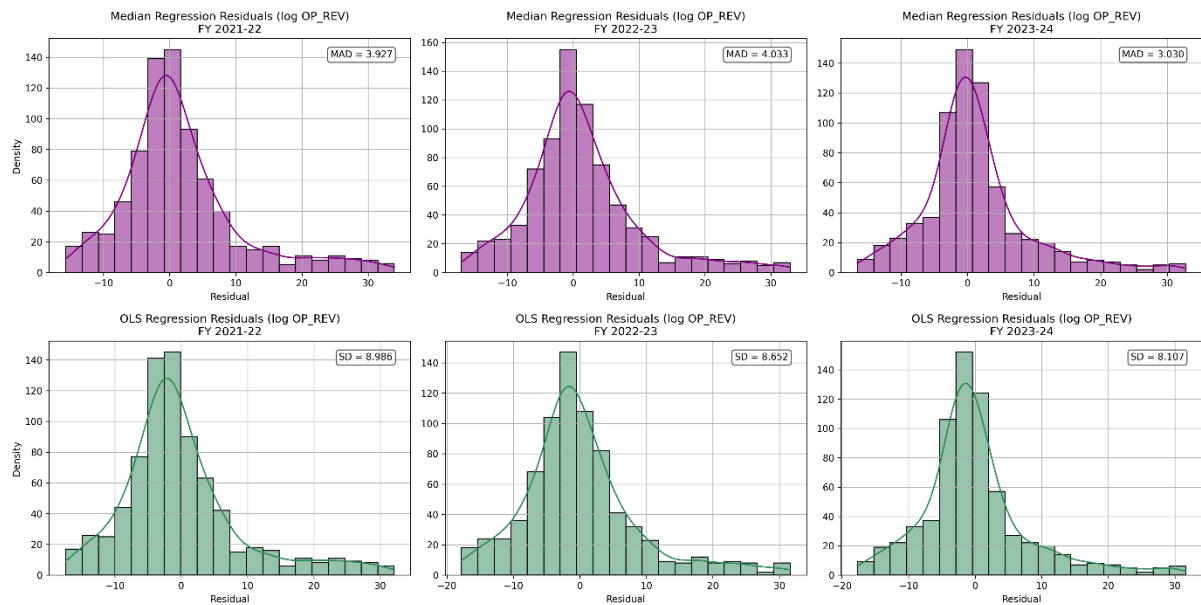
OLS Regression of PLI vs log(Op\_Rev) by FY



Median Regression of PLI vs Op\_Rev (log scale) by FY



#### Residual Distributions: Median Regression (MAD) vs OLS (SD) using log(OP\_REV) by FY



#### Interpretation:

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

##### OLS regression

- $R^2$  values are negligible ( $<0.003$ ).
- Slopes are statistically insignificant across all years (p-values 0.17–0.83).
- **Inference:** No meaningful association between operating revenue and profitability for export-oriented firms.

##### Median regression

- Slopes are extremely small and non-significant (p-values 0.41–0.91).
- Pseudo- $R^2$  values are essentially zero.
- **Inference:** The direct OR–PLI relationship vanishes entirely in this export-focused cohort.

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

##### OLS regression

- FY 2021-22 shows a weakly significant slope ( $p \approx 0.04$ ,  $R^2 \approx 0.005$ ).
- FY 2022-23 and 2023-24 are statistically insignificant (p-values 0.64–0.95).
- **Inference:** OLS provides inconsistent and unstable results.

##### Median regression

- FY 2021-22: Positive, significant slope ( $p \approx 0.0017$ ) with modest pseudo- $R^2$  (0.006).

- FY 2022-23: Positive, weaker but still significant slope ( $p \approx 0.022$ ).
- FY 2023-24: No relationship at all (slope  $\approx 0$ ,  $p \approx 1.0$ ).
- Residual MAD ( $\approx 3.0$ – $4.0$ ) is consistently much lower than OLS residual SD ( $\approx 8.1$ – $9.0$ ), showing better stability.
- **Inference:** Median regression with  $\log(\text{OR})$  captures some positive size–profitability link in the first two years, but the effect disappears completely in the most recent year.

---

### Overall takeaway for Scenario 3

- Export-oriented firms show **no reliable relationship** between direct OR and profitability.
- Median regression with  $\log(\text{OR})$  offers **partial explanatory power** in earlier years, but the relationship breaks down in FY 2023-24.

---

### Prediction Table:

Based on the analysis, the median regression with  $\log(\text{OR})$  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).

FY	2021-22	2022-23	2023-24	Average
<b>OP_REV</b>				
<b>1</b>	14.53	15.36	16.84	15.58
<b>100</b>	16.34	16.74	16.84	16.64
<b>300</b>	16.77	17.07	16.84	16.89
<b>500</b>	16.97	17.23	16.84	17.01
<b>1000</b>	17.24	17.43	16.84	17.17
<b>1500</b>	17.40	17.56	16.84	17.27
<b>2000</b>	17.51	17.64	16.84	17.33
<b>3000</b>	17.67	17.76	16.84	17.42
<b>5000</b>	17.87	17.92	16.84	17.54
<b>10000</b>	18.14	18.12	16.84	17.70
<b>20000</b>	18.42	18.33	16.84	17.86

## **Scenario 4 – Export Oriented & Uncontrolled**

### **Scenario description**

This scenario applies the same filters as Scenario 3 (export filter + removal of outliers) but adds an **RPT filter**, restricting the sample to uncontrolled firms only. The applied filters are:

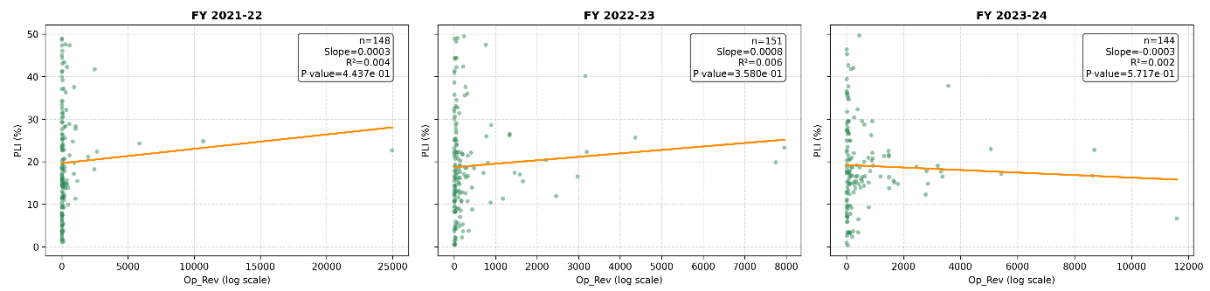
- **Net worth > 0**
- **Persistent loss = 0**
- **Employee cost  $\geq$  25% of OR**
- **Service income  $\geq$  75% of OR**
- **Operating revenue between 1 and 30,000**
- **PLI between 0 and 50**
- **Export income  $\geq$  75% of OR**
- **RPT\_<25%**

Final sample sizes: **FY 2021-22: 148, FY 2022-23: 151, FY 2023-24: 144** firm-year observations.

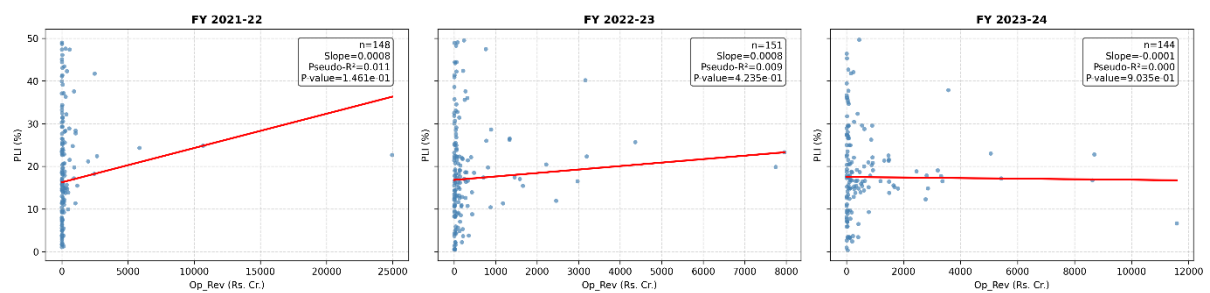
---

## Results:

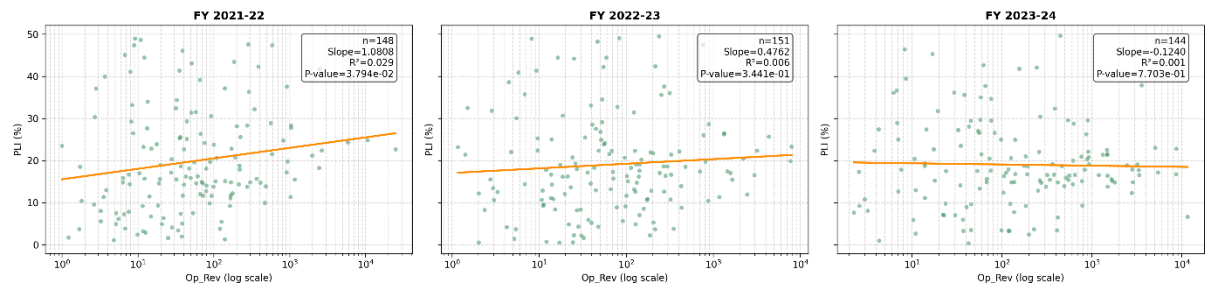
OLS Regression of PLI vs Op\_Rev by FY



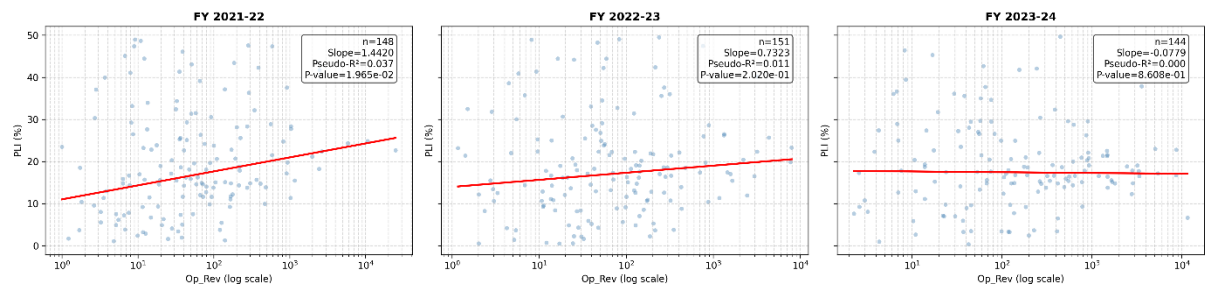
Median Regression of PLI vs Op\_Rev by FY



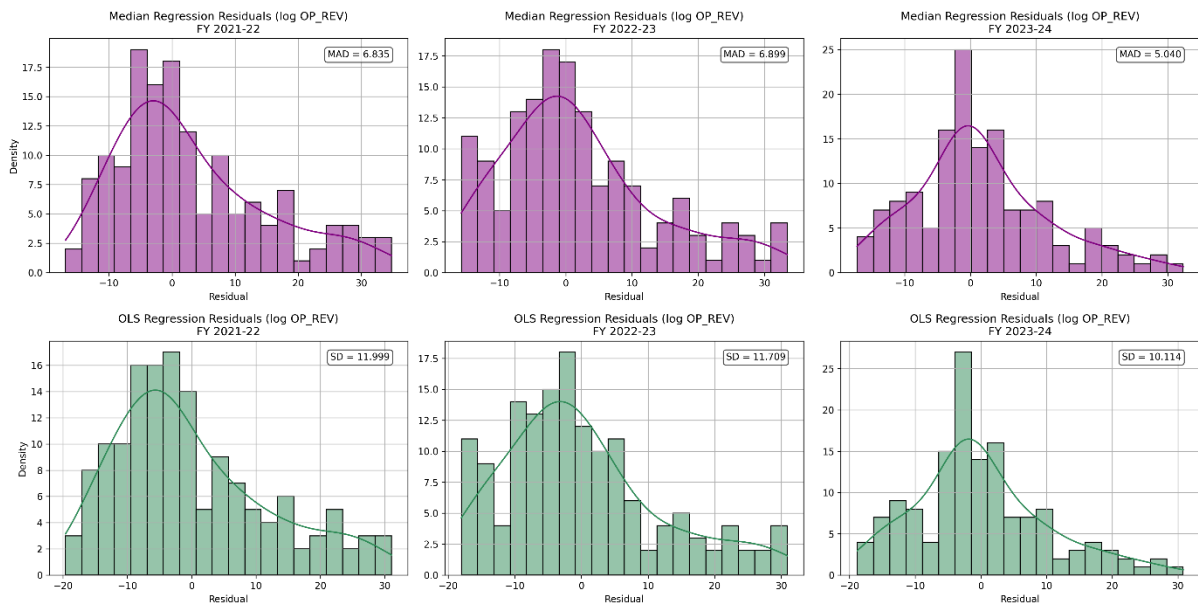
OLS Regression of PLI vs log(Op\_Rev) by FY



Median Regression of PLI vs Op\_Rev (log scale) by FY



### Residual Distributions: Median Regression (MAD) vs OLS (SD) using log(OP\_REV) by FY



### Interpretation:

#### 1. Sample Size

- Very small dataset ( $n \approx 144\text{--}151$  per FY), compared to 1,400+ in Scenario-2.
- Indicates only a limited set of truly comparable uncontrolled, export-oriented IT/ITeS companies survive after strict filters.
- Small  $n \rightarrow$  higher sensitivity of slope estimates, noisier results.

#### 2. PLI vs Operating Revenue (linear scale)

- **Median regression slopes:** small, positive in 2021-22 & 2022-23, but statistically insignificant; negative in 2023-24.
- **OLS slopes:** similar direction, none significant.
- **Takeaway**  $\rightarrow$  On raw scale, no meaningful linear relationship between size (op\_rev) and profitability.

#### 3. PLI vs Log(Operating Revenue)

- **2021-22:**
  - Median regression slope = **1.44** ( $p = 0.0196$ , pseudo- $R^2 \approx 0.037$ )  $\rightarrow$  significant.
  - OLS slope = **1.08** ( $p = 0.038$ ,  $R^2 \approx 0.029$ )  $\rightarrow$  significant.
  - Interpretation: **positive size effect** – larger export-oriented uncontrolled companies earned higher PLI.

- **2022-23 & 2023-24:**
    - Slopes drop and lose significance ( $p > 0.2$ ,  $p > 0.8$ ).
    - Suggests the 2021-22 effect did not persist consistently across years.
- 

#### 4. Residuals (Fit Quality)

- Median regression MAD: ~5–7
  - OLS residual SD: ~10–12
  - Again, median regression gives **tighter residuals** → more robust fit against outliers even in small sample.
- 

#### Interpretation for Policy / TP Safe Harbour Context

- When filtering strictly to **export-oriented, uncontrolled companies**,
    - The relationship between firm size and profitability is **weak or unstable** across years.
    - A statistically significant positive slope appears in **2021-22**, but **not robust across 2022-23 and 2023-24**.
    - This suggests that **size is not a reliable driver of PLI in this narrowly defined uncontrolled set**.
  - Median regression provides more consistent residual control, but even then explanatory power is low ( $\text{pseudo-}R^2 \leq 0.04$ ).
- 

#### Overall Takeaway:

Scenario-4 confirms that once you restrict to **pure export-oriented uncontrolled comparables**, the **size-profitability link largely disappears** (except one year blip). This makes a strong case that **PLI ranges should not be adjusted for size effects under such strict comparability filters**, since the relationship is not systematic across years.

---



**Prediction Table:**

<b>FY</b>	<b>2021-22</b>	<b>2022-23</b>	<b>2023-24</b>	<b>Average</b>
<b>OP_REV</b>				
<b>1</b>	11.00	13.95	17.81	14.25
<b>100</b>	17.64	17.32	17.45	17.47
<b>300</b>	19.22	18.13	17.37	18.24
<b>500</b>	19.96	18.50	17.33	18.60
<b>1000</b>	20.96	19.01	17.27	19.08
<b>1500</b>	21.55	19.31	17.24	19.37
<b>2000</b>	21.96	19.52	17.22	19.57
<b>3000</b>	22.55	19.81	17.19	19.85
<b>5000</b>	23.28	20.19	17.15	20.21
<b>10000</b>	24.28	20.69	17.09	20.69
<b>20000</b>	25.28	21.20	17.04	21.17

## Overall Takeaways Across Scenarios

### 1. Effect of Filtering on Sample Size

- The dataset shrinks from ~1,500 entities in Scenario-1 (No Filters) to barely ~150 in Scenario-4 (All standard TPO filters) as stricter filters (export intensity, RPT thresholds) are applied.
- While these filters improve comparability, they also reduce statistical power. Thus, results for Scenarios 3–4 must be read with caution, since limited observations constrain statistical inference.

### 2. PLI vs OR (Linear Form)

- Across all scenarios, direct linear regression ( $PLI \sim OR$ ) produced very low  $R^2$  and insignificant slopes.
- Conclusion: **Raw operating revenue is not a reliable predictor of profitability.**

### 3. PLI vs Log(OR) (Functional Transformation)

- In broader datasets (Scenarios 1–2), **median regression slopes are consistently positive and statistically significant across years** → strong evidence of a **positive size–profitability effect**.
- **Scenario 2 (outliers removed, only service income, employee cost filters, no export/rpt filters)** is the most representative dataset of the IT/ITES industry. Median regression in this case gives most meaningful results too. slopes are in range of **0.5–0.6**, meaning:
  - A **doubling of revenue** is associated with a **0.4–0.45 percentage point increase in PLI**.
  - Put differently, it takes about a **170% increase in revenue to achieve a 1% absolute increase in PLI**.
- This shows that the effect is **statistically robust but economically modest**.
- In stricter datasets (Scenarios 3–4), the relationship weakens, but this is largely attributable to much smaller sample sizes.
- Overall → **Log transformation is essential** for capturing size effects.

### 4. OLS vs Median Regression

- In every scenario, **median regression delivers tighter residuals (lower MAD)** than OLS, highlighting its robustness against outliers and skewness.
- **Median regression on log(OR)** is therefore the most preferred model.

---

## Final Interpretation and Policy Implications

- The **best model** is **median regression of PLI on log(OR)**.
- **Scenario 2 (outliers removed, broad set)** provides the most reliable evidence: a **positive and consistent relationship across years**, but of **modest magnitude**.

- This indicates that while larger IT/ITeS firms tend to achieve slightly higher margins, the effect is too gradual to justify **hard revenue thresholds** as an eligibility criterion.
- Instead, the findings support a shift towards a **tiered safe harbour margin structure**:
  - Acceptable margins could be set **slightly lower for smaller firms and slightly higher for larger firms**, aligning with the observed data.
  - This approach removes rigid cut-offs, reflects economic reality more accurately, and balances simplicity with fairness in compliance.