**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 ~ OR) produced very low R² 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.