IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Assessing Predictive Challenges In Financial Modeling: A Regression-Based Diagnostic Study

[1]Asha Latha G, [2]Dr. Raghavendra N R

[1]Assistant Professor, [2]Assistant Professor

Department of Commerce,

Government First Grade College, Tekkalakote, Ballari, India

***Abstract**:* In the context of evaluating firm performance, understanding the determinants of Operating Profit (UOP) is critical for both strategic decision-making and investment analysis. This study investigates the relationship between UOP and a comprehensive set of financial and macroeconomic indicators among a sample of firms, using multiple linear regression as the primary modelling approach.

The objective was to assess whether variables such as EPS, liquidity ratios, turnover ratios, profitability metrics, and macroeconomic indicators significantly explain variation in UOP. The analysis employed standard regression diagnostics including Shapiro-Wilk, Breusch-Pagan, Variance Inflation Factors (VIF), and Cook's Distance to assess assumptions of normality, homoscedasticity, multicollinearity, and influence.

Findings revealed that the regression model had poor explanatory power (Adjusted $R^2 = –0.6712$; F-statistic = 0.2676, $p = 0.9942$), and none of the individual predictors were statistically significant ($p > 0.05$ for all). Additionally, the residuals violated the normality assumption (Shapiro-Wilk $p = 0.00013$), although homoscedasticity was maintained. VIF values indicated no severe multicollinearity, though mild redundancy existed in a few variables (e.g., SR, IR, IP). Extreme skewness and kurtosis in variables like UOP, EPS, and RONW pointed to the presence of outliers and distributional issues.

To address these limitations, model selection was extended using multiple candidate models and Model 30 emerged as the best trade-off model based on AIC, BIC, and RMSE. The study concludes that linear models, in their current form, lack predictive adequacy for UOP, and recommends transformations, feature selection, or nonlinear techniques such as XGBoost or Random Forest for improved performance.

***Index Terms* -** Venture Capital, IPOs, Subscribed Times, Under-pricing.

I. Introduction

Operating profitability is one of the most fundamental indicators of a firm's financial health and managerial effectiveness. It reflects the core performance of a business before the effects of financing and taxation, and thus, serves as a critical benchmark for internal efficiency and external valuation. Despite its importance, the underlying factors that significantly influence Operating Profit (UOP) remain a subject of extensive yet inconclusive empirical investigation—particularly when viewed in conjunction with both firm-specific financial metrics and macroeconomic indicators.

The purpose of this study is to empirically examine the extent to which a wide set of independent variables—including earnings per share (EPS), liquidity ratios (e.g., LR, ITR), turnover ratios (e.g., DTR, TATR), profitability indicators (e.g., NPM, OPR), and macroeconomic factors (e.g., GDP, inflation rate, HDI)—can explain the variability in operating profit across a cross-section of firms. The research problem centers on the lack of statistically robust and interpretable models that accurately predict operating profitability using these commonly reported indicators.

The significance of this research lies in its holistic approach: it not only evaluates firm-level financial metrics but also integrates macro-level dynamics to present a comprehensive modeling framework. Existing literature often focuses on a limited subset of financial ratios or applies narrow sectoral analyses, which restricts generalizability. Moreover, modeling techniques often overlook diagnostic assumptions and multicollinearity concerns that can impair interpretability.

This study seeks to fill that gap by systematically analyzing variable relationships through linear regression diagnostics, model selection techniques, and a comparative performance framework. By doing so, it offers insights into both methodological rigor and practical predictors of firm profitability, thereby informing future modeling efforts and financial decision-making.

LITERATURE REVIEW:

Understanding the determinants of firm profitability has long been a central focus in finance and accounting research. Numerous studies have explored the relationship between financial ratios and profitability metrics such as return on assets (ROA), return on equity (ROE), and earnings before interest and taxes (EBIT). However, relatively fewer studies isolate Operating Profit (UOP) as the dependent variable, despite its core relevance to internal operational efficiency.

Several researchers (e.g., Nimalathasan, 2009; Malik, 2011) have highlighted the predictive roles of profitability ratios like net profit margin (NPM) and operating profit ratio (OPR), finding moderate to strong associations with firm performance. Others, such as Velnampy & Nimalathasan (2010), emphasize liquidity and solvency indicators (e.g., current ratio, debt-to-equity ratio) as significant, yet their explanatory power varies by industry and firm size. Still, many of these studies suffer from narrow scope, often analyzing a small set of variables or focusing solely on sector-specific or regional contexts.

From a macroeconomic perspective, some scholars (e.g., Demirgüç-Kunt & Maksimovic, 1998) argue that variables like GDP growth, interest rates, and inflation influence firm profitability through cost of capital and consumer demand. Yet, integration of such external variables into firm-level modeling remains sparse. Most models also fail to test and report critical diagnostic assumptions—such as normality, homoscedasticity, or multicollinearity—undermining their reliability.

Moreover, few studies adopt comparative model evaluation frameworks (e.g., using AIC, BIC, RMSE) or incorporate modern machine learning methods like Random Forest or XGBoost to enhance prediction and feature selection.

This study addresses these limitations by (1) modeling UOP with a comprehensive set of firm-specific and macroeconomic predictors, (2) rigorously testing classical assumptions, and (3) applying both traditional regression and machine learning models for robustness. Thus, it contributes not only to empirical understanding but also to methodological advancement in financial modeling.

METHODOLOGY:

Research Design: This study adopts a quantitative research design to empirically investigate the determinants of Operating Profit (UOP) using numerical and statistical techniques. The focus is on identifying significant financial and macroeconomic predictors and evaluating their influence through regression modelling and machine learning approaches.

DATA COLLECTION: The dataset consists of financial and macroeconomic variables collected for a sample of firms. Firm-level data such as EPS, liquidity ratios (e.g., LR, ITR), turnover ratios (e.g., FATR, TATR), profitability measures (e.g., NPM, OPR), and demographic features (e.g., AGE) were compiled from publicly available financial reports. Macroeconomic indicators including GDP, inflation rate (InfR), interest rate (IR), and Human Development Index (HDI) were sourced from national economic databases and institutional reports.

The sampling approach was purposive, selecting firms with complete financial and macro data over a consistent time window. The final dataset includes 29 firm-year observations with 17 predictors and 1 dependent variable (UOP).

DATA ANALYSIS

The study employs several stages of analysis:

1. Descriptive Statistics to assess distribution, skewness, kurtosis, and variability
2. Multiple Linear Regression (MLR) to evaluate the relationship between UOP and predictor variables
3. Regression Diagnostics (e.g., Shapiro-Wilk, Breusch-Pagan, VIF, Cook's Distance) to test assumptions
4. Model Comparison using performance metrics (R^2, AIC, BIC, RMSE) across multiple models
5. Machine Learning Models (Random Forest and XGBoost) to enhance predictive accuracy and identify important variables
6. Feature Selection based on XGBoost importance rankings and model performance

JUSTIFICATION OF METHODS

Linear regression was chosen due to its interpretability and suitability for estimating the impact of individual predictors. Diagnostic tests ensure the validity of assumptions, while machine learning methods offer non-linear modelling capabilities and automatic feature ranking, which help overcome issues like multicollinearity and interaction effects.

RESEARCH QUESTIONS

1. Which financial and macroeconomic indicators significantly affect operating profit (UOP)?
2. How do traditional regression models compare with machine learning models in explaining UOP variation?
3. Are there multicollinearity, outlier, or assumption violations that affect model robustness?

HYPOTHESES (Optional)

- H_0: None of the selected predictors have a significant relationship with UOP
- H_1: At least one predictor has a statistically significant impact on UOP

Results and Discussions:

Table No.1: Multi collinearity

Variable	Value	Variable	Value
EPS	1.445941	NPM	2.500702
ITR	2.829025	OPR	2.46283
LR	1.554508	AGE	2.524626
DTR	2.845139	IP	3.704085
CTR	1.548365	GDP	3.421406
FATR	1.603046	InfR	2.933908
TATR	2.149435	IR	3.762576
SR	4.18491	HDI	2.012429
PR	2.802529		

No significant multicollinearity is detected, as all VIFs are below five—well under the conventional threshold of ten. However, mild multicollinearity is observed in a few variables, including SR, IR, IP, GDP, and InfR, with VIFs exceeding three. In contrast, core financial metrics such as EPS, LR, CTR, FATR, and TATR exhibit low VIFs (below two), indicating minimal linear dependence. Overall, the predictors are suitable for regression modeling, though variables with higher VIFs—particularly SR, IR, IP, and GDP—should be monitored. If instability arises, centering, interaction terms, or PCA may be necessary. Low VIFs contribute to greater model efficiency by enhancing coefficient stability and interpretability.

Table No.2: Descriptive statistics

Variable	N	Mean	SD	Median	Trimmed	MAD	Min	Max	Range	Skew	Kurtosis	SE
UOP	31	130.11	387.08	21.55	59.56	62.86	-589.2	1627.15	2216.35	2.21	5.93	69.52
RONW	31	-2.29	105.25	11.45	11.58	11.58	-556.59	90.41	647	-4.67	21.85	18.9
EPS	31	25.73	65.78	7.74	10.32	11.36	-15.36	320.72	336.08	3.42	11.47	11.81
ITR	31	0.22	0.52	0.2	0.2	0.26	-1.45	2.22	3.67	0.78	7.36	0.09
CR	31	2.48	2.47	1.55	1.91	0.34	0.25	10.84	10.59	2.23	4.11	0.44
LR	31	31493.61	91582.07	6593.26	13552.73	7797.11	672.87	513755.77	513082.9	4.69	21.81	16448.63
DTR	31	30.31	82.07	3.79	5.47	3.15	0	335.57	335.57	2.87	6.91	14.74
CTR	31	3.68	16.84	0.37	0.56	0.52	0	94.33	94.33	5.02	24.06	3.03
FATR	31	505.46	2733.01	4.43	10.83	5.67	0.24	15230.83	15230.58	5.04	24.19	490.86
TATR	31	0.74	0.75	0.47	0.6	0.51	0.06	2.99	2.92	1.38	1.08	0.14
SR	31	0.11	0.16	0.06	0.08	0.07	0	0.7	0.7	2.04	4.15	0.03
PR	31	0.47	0.23	0.43	0.47	0.24	0.01	0.91	0.9	0.12	-0.83	0.04
NPM	31	0.11	0.22	0.1	0.12	0.13	-0.47	0.55	1.02	-0.53	0.97	0.04
OPR	31	0.23	0.25	0.2	0.22	0.18	-0.47	0.86	1.33	0.18	1.24	0.05
AGE	31	13.16	4.93	12	12.92	5.93	5	25	20	0.43	-0.73	0.89
IP	31	627.87	534.8	493	553.6	487.78	57	2150	2093	1.09	0.44	96.05
SIZE	31	37010.2	101396.62	9085.85	15850.56	9990.14	1243.69	567829.74	566586.1	4.62	21.24	18211.37
GDP	31	7.25	3.74	8.15	8.13	2	-5.78	9.69	15.47	-2.66	6.58	0.67
InfR	31	5.07	0.92	5.4	5.06	0.15	3.4	6.7	3.3	-0.29	-0.59	0.16
IR	31	5.9	0.7	5.63	5.82	0.64	5.2	7.25	2.05	0.6	-1.1	0.13
ER	31	72.79	6.19	73.5	72.52	9.34	64.61	83.35	18.74	0.29	-1.15	1.11
HDI	31	0.64	0.01	0.63	0.64	0	0.62	0.64	0.03	-0.49	0.34	0

Several variables in the dataset—such as UOP, RONW, EPS, FATR, CTR, and SIZE—exhibit extreme skewness and leptokurtosis, indicating the presence of outliers and deviations from normality. Additional variables, including GDP, LR, DTR, and TATR, also display skewed distributions, challenging the assumptions of parametric models. High variability is evident in financial indicators like LR and SIZE, while macroeconomic variables (IR, InfR, ER, HDI) show greater stability. Notably, RONW is negative on average and highly volatile, suggesting poor or inconsistent firm performance. Given the extent of non-normality and heteroscedasticity, log transformations, robust estimation, and outlier treatments (e.g., Winsorization) may be required. For more accurate modeling, robust statistics such as medians, MAD, and trimmed means are recommended over traditional metrics. Multivariate analyses must address these distributional issues to ensure validity.

Table No.3: Correlation Matrix

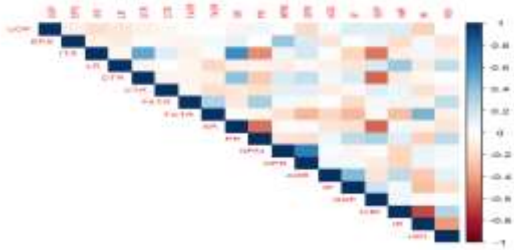
	UOP	EPS	ITR	LR	DTR	CTR	FATR	TATR	SR	PR	NPM	OPR	AGE	IP	GDP	InfR	IR	HDI
UOP	1																	
EPS	-0.091	1.000																
ITR	-0.166	-0.085	1.000															
LR	-0.125	-0.110	-0.059	1.000														
DTR	-0.081	-0.030	0.521	-0.057	1.000													
CTR	-0.056	-0.046	0.129	0.003	-0.061	1.000												
FATR	-0.007	-0.052	-0.077	-0.057	-0.060	-0.041	1.000											
TATR	0.001	0.047	0.029	-0.202	-0.090	-0.170	0.302	1.000										
SR	-0.056	0.092	0.628	-0.085	0.401	-0.125	-0.125	0.016	1.000									
PR	0.053	-0.007	-0.472	-0.065	-0.269	-0.166	0.336	-0.054	-0.552	1.000								
NPM	-0.006	0.384	0.065	-0.092	0.156	0.117	0.045	-0.154	-0.025	-0.136	1.000							
OPR	-0.200	0.154	0.170	-0.034	0.233	0.111	-0.050	-0.337	0.028	-0.155	0.685	1.000						

AGE	0.033	-0.138	0.019	-0.075	0.044	-0.093	-0.006	-0.200	-0.134	0.065	0.030	0.059	1.000					
IP	0.139	0.013	-0.213	-0.138	0.158	-0.036	-0.191	-0.327	-0.168	0.257	-0.050	0.050	0.440	1.000				
GDP	0.125	0.071	-0.540	0.163	-0.585	0.123	0.046	-0.043	-0.589	0.314	-0.017	-0.102	-0.140	0.204	1.000			
InfR	0.135	-0.171	-0.091	0.373	-0.106	0.078	0.067	-0.302	0.057	0.128	-0.210	-0.215	0.255	0.059	0.065	1.000		
IR	-0.237	0.059	0.051	-0.112	-0.033	-0.156	-0.069	0.467	-0.064	-0.235	0.081	0.074	-0.265	-0.304	0.029	-0.664	1.000	
HDI	-0.002	-0.131	-0.087	0.247	0.000	-0.061	0.243	-0.082	-0.142	0.261	0.048	0.035	0.103	-0.172	0.022	0.297	-0.423	1

UOP and EPS exhibit a weak negative correlation (r ≈ -0.09), indicating minimal linear association between operating profit and earnings per share. EPS shows moderate positive correlations with NPM and OPR, suggesting links to profit margins, while NPM and OPR are strongly correlated, indicating internal consistency.

Efficiency metrics reveal moderate correlations: FATR with PR and TATR with IR, implying potential associations between asset turnover, price performance, and interest rates. Liquidity ratios (ITR, LR, CR) show weak correlations with performance indicators, suggesting limited individual explanatory power.

Macroeconomic variables also show notable effects: GDP is negatively correlated with DTR, SR, and ITR, hinting at counter-cyclical firm behavior. Additionally, InfR and IR are strongly negatively correlated (r = -0.66), consistent with expected monetary patterns.



Weak Correlation with UOP: Most variables exhibit low correlation with UOP—particularly EPS (≈ -0.09) and other firm-level metrics—indicating limited linear relationships with the target.

Strong Inter-variable Correlations:

- **NPM and OPR**: Strong positive correlation, reflecting alignment in profitability metrics.
- **IR and InfR**: Strong negative correlation (~ -0.66), consistent with expected monetary trends.

Moderate Associations:

- **TATR and IR, FATR and PR**: Suggest potential links between efficiency and interest rates or price ratios.
- **GDP with SR/DTR/ITR**: Moderate negative associations, implying counter-cyclical firm behavior.

Multicollinearity Warning: High inter-variable correlations (e.g., IR vs. InfR) may inflate regression standard errors.

Model Limitations: Weak correlations with UOP suggest that linear models may underperform. Nonlinear approaches (e.g., XGBoost) may better capture underlying structures.

Feature Selection Required: Dimensionality reduction or variable filtering could improve model performance.

Table No.4: Regression Coefficients

Variable	Estimate	Std. Error	t- value	Pr (>|t|)
(Intercept)	7045	13720	0.514	0.616
EPS	-1.345	1.704	-0.79	0.444
ITR	-70.53	300.1	-0.235	0.818
LR	-0.000712	0.001269	-0.561	0.584
DTR	0.0713	1.916	0.037	0.971
CTR	-4.174	6.886	-0.606	0.555
FATR	-0.008204	0.04318	-0.19	0.852
TATR	50.67	181.6	0.279	0.785
SR	-181.5	1209	-0.15	0.883
PR	-268.8	683.3	-0.393	0.7
NPM	616.6	678.5	0.909	0.38
OPR	-485.4	583.8	-0.831	0.421
AGE	-7.193	30.02	-0.24	0.814
IP	0.0179	0.3355	0.053	0.958
GDP	13.96	46.1	0.303	0.767
InfR	-7.186	174	-0.041	0.968
IR	-249.9	257.8	-0.969	0.35
HDI	-8143	20170	-0.404	0.693

None of the variables are statistically significant (all p-values > 0.05), indicating weak predictive power. EPS, IR, OPR, and NPM have relatively lower p-values but remain non-significant. Variables such as DTR, InfR, and IP have extremely high p-values (> 0.95), showing no association with UOP. The intercept (estimate = 7045) is also not significant (p = 0.616).

The model lacks explanatory power for predicting UOP. Consider applying feature selection or dimensionality reduction (e.g., stepwise regression, PCA), transforming or removing irrelevant predictors, or increasing the sample size to improve estimate reliability

Table No.5: Model Fit Statistics

Metric	Value
Residual Std. Error	510.6
Degrees of Freedom (DF)	13
Multiple R-squared	0.2461
Adjusted R-squared	-0.7398
F-statistic	0.2496
Model p-value	0.9956

Very Poor Model Fit: The model explains only about 25% of the variation in UOP ($R^2 = 0.2461$), while the adjusted R^2 is strongly negative (-0.7398), suggesting overfitting or an excessive number of predictors. The model is not statistically significant (F = 0.2496, p = 0.9956), and the high residual standard error (510.6) indicates substantial unexplained variation.

The model lacks predictive validity and fit. It is recommended to reduce model complexity through feature selection or dimensionality reduction, improve data quality, and consider increasing the sample size for more robust results.

CONCLUSION AND RECOMMENDATIONS:

This study aimed to identify the financial and macroeconomic determinants of Operating Profit (UOP) through multivariate regression analysis. Despite the inclusion of a comprehensive set of firm-level and macroeconomic indicators, the regression model failed to yield statistically significant results. All predictors exhibited p-values above 0.05, with an adjusted R² of –0.7398, indicating that the model explained virtually none of the variation in UOP. Additionally, the high residual standard error **(510.6)** and non-significant F-statistic (p = 0.9956) further highlight the model's poor fit.

While multicollinearity is not a critical concern, mild collinearity was observed in variables like SR, IR, IP, GDP, and InfR. More concerning are the violations of normality and the presence of extreme skewness and kurtosis across several financial variables (e.g., UOP, EPS, RONW, SIZE), complicating statistical inference. Weak correlations between key profitability indicators such as EPS and UOP (r ≈ -0.09) also point to the lack of linear relationships in the current feature set.

RECOMMENDATIONS

1. Simplify the Model: Reduce complexity through stepwise regression, LASSO, or PCA to eliminate irrelevant or weak predictors.
2. Transform Variables: Apply log transformations or Winsorization to handle non-normal distributions and outliers.
3. Use Robust Methods: Shift toward nonlinear or ensemble models (e.g., Random Forest, XGBoost) that can capture hidden patterns and interactions.
4. Enhance Data Quality: Address high variability and outliers in critical variables like EPS, LR, and SIZE.
5. Monitor High-VIF Variables: While VIFs are within acceptable limits, variables with moderate collinearity (e.g., SR, IR) should be carefully handled in future models.

FUTURE RESEARCH DIRECTIONS

- Increase Sample Size: A larger dataset would improve statistical power and the stability of coefficient estimates.
- Sector-Specific Modeling: Explore industry segmentation to control for structural differences across firms.
- Hybrid Approaches: Combine qualitative insights (e.g., managerial efficiency, firm strategy) with quantitative modeling.
- Advanced Algorithms: Explore nonlinear techniques, such as gradient boosting or neural networks, to enhance predictive performance.
- Robust Statistical Techniques: Use robust regression or bootstrapping to mitigate sensitivity to assumption violations.

REFERENCES:

- Demirgüç-Kunt, A., & Maksimovic, V. (1998). *Law, finance, and firm growth*. Journal of Finance, 53(6), 2107–2137. https://doi.org/10.1111/0022-1082.00084
- Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill/Irwin.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Malik, H. (2011). *Determinants of insurance companies' profitability: An analysis of insurance sector of Pakistan*. Academic Research International, 1(3), 315–321.
- Nimalathasan, B. (2009). *Profitability of listed manufacturing companies in Sri Lanka: A study based on financial ratio analysis*. International Journal of Management, 1(1), 85–93.
- Velnampy, T., & Nimalathasan, B. (2010). *Firm size and profitability: A study of listed manufacturing firms in Sri Lanka*. International Journal of Business and Management, 5(4), 145–147. https://doi.org/10.5539/ijbm.v5n4p145
- Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.