APPLICATION OF LINEAR REGRESSION MODEL IN FORECASTING CORPORATE INCOME TAX REVENUE: A CASE STUDY OF TAX SUB-DEPARTMENT REGION I

Assoc.Prof.PhD. Truong Thi Thuy* - PhD. Truong Thi Thu Ha**

Abstract: ***This study analyzes the applicability of the multiple linear regression model in forecasting corporate income tax revenues based on firm-level financial performance indicators. Utilizing data from 213 enterprises under the jurisdiction of Regional Tax Office I during the 2023-2024 period, the research examines the influence of variables such as cost of goods sold (COGS), liquidity ratio, and asset utilization efficiency on return on sales (ROS), which is used as a proxy for financial performance. The regression results indicate that these variables significantly explain the variation in ROS, thereby reflecting a correlation with corporate income tax obligations.***

• Keywords: *corporate income tax, revenue forecasting, financial performance, linear regression, tax administration.*

Date of receipt: 13th Feb., 2025
Date of delivery revision: 20th Mar., 2025
Date of receipt revision: 28th May, 2025
Date of approval: 02nd Jul., 2025
DOI: https://doi.org/10.71374/jfar.v25.i4.06

1. Introduction

Forecasting corporate tax revenues plays a crucial role in shaping fiscal policy and public budget planning. In many developing economies particularly at the local level tax authorities often face difficulties in constructing accurate forecasting models due to the lack of dynamic tools tailored to the characteristics of local enterprises. While macroeconomic indicators provide a top-down perspective, they often fail to capture micro-level factors such as the financial health or operational efficiency of individual firms.

Traditional tax forecasting methods typically rely on historical trends or fixed growth assumptions, which can lead to misestimations of actual fiscal performance. As a result, budget allocation and compliance planning may be adversely affected. In this context, adopting a quantitative approach specifically regression analysis to forecast tax revenues based on firms' financial indicators emerges as a potentially viable solution.

2. Research Objectives

To develop and validate a multiple linear regression model for estimating the return on sales (ROS) based on key financial indicators.

To identify the relationships between financial performance and variables such as firm size, financial leverage, liquidity, tangible assets, cost of goods sold (COGS), and asset utilization efficiency.

To apply the model for forecasting corporate income tax obligations and enhancing the accuracy of public revenue projections.

This study contributes to the body of tax analysis literature by proposing a replicable quantitative model, particularly useful for local tax authorities. It aims to improve forecasting precision and support risk management in tax administration.

3. Data and Research Methodology

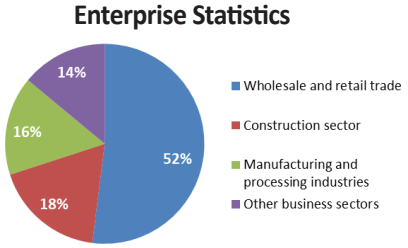
3.1. Research Sample

The dataset comprises financial statements and corporate tax declarations from 213 enterprises for the period 2023-2024, extracted from the centralized taxpayer database. The selection of enterprises was based on the following criteria:

Availability of complete financial reports and tax declarations for three consecutive years.

The industry distribution of the sample is as follows:

+ 67% operate in wholesale and retail trade,
+ 15% in the construction sector,
+ 13% in manufacturing and processing industries,
+ 5% in other business sectors.



3.2. Research Model

The author employs STATA software version 17 to

* *Academy of Finance*
** *Cau Giay district Tax Team*

estimate the regression equations.

A multiple linear regression model is constructed to examine the impact of corporate financial variables on financial performance. The model includes the following variables:

ROS: Return on Sales

SIZE: Firm size (logarithm of total assets)

LIQ: Liquidity ratio (logarithm of the liquidity ratio)

LEV: Financial leverage ratio (logarithm of financial leverage)

TANG: Tangible asset ratio

COGS: Cost of Goods Sold as a percentage of revenue

AUT: Asset utilization ratio (revenue / total assets)

The linear regression model is applied to test the linear relationship between financial performance and influencing factors, including financial health and operational efficiency. Linear regression is a data analysis technique used to predict the value of a dependent variable based on the known values of one or more independent variables.

The linear regression equation is specified as follows:

$$ROS = \beta_0 + \beta_1 SIZE + \beta_2 LIQ + \beta_3 LEV + \beta_4 TANG + \beta_5 COGS + \beta_6 AUT + \varepsilon$$

Dependent and Independent Variables

The dependent variable (i.e., the variable being influenced) is financial performance, measured by Return on Sales (ROS). Financial performance is linearly related to the amount of corporate income tax payable. This means that, assuming other factors remain constant, a one-unit increase in financial performance will correspond to an increase in corporate income tax obligations.

The independent variables (i.e., influencing variables) are grouped into two categories of indicators:

Financial Health Indicators

These serve as fundamental factors in evaluating a firm's ability to meet financial commitments, withstand economic downturns, and sustain ongoing operations. The group includes:

Firm Size (SIZE): Measured by the total assets (both current and non-current assets) of the enterprise.

Liquidity Ratio (LIQ): Calculated as the ratio of total current assets to total short-term liabilities. A higher liquidity ratio enhances the firm's ability to overcome financial challenges and focus on profit growth, while a lower ratio indicates vulnerability.

Financial Leverage (LEV): Defined as the ratio of total debt to total assets. This indicator assesses the level of financial risk associated with using debt to support business activities and investment. While increased leverage can positively impact financial performance, excessive leverage may raise financial costs, ultimately impairing the firm's performance.

Asset Tangibility (TANG): Represents the ratio of fixed assets to total assets. A higher asset tangibility is generally favorable as it implies lower reliance on external financing, which in turn positively affects financial performance.

Operational Efficiency Indicators

These are related to a firm's ability to utilize its resources efficiently to generate revenue and control costs:

Cost of Goods Sold (COGS): Defined as the ratio of direct production and service delivery costs to total revenue. COGS has a correlational relationship with financial performance; higher production costs may reflect increased production volumes, potentially leading to greater sales and revenue.

Asset Utilization (AUT): Measured by the ratio of revenue to total assets. This indicator reflects a firm's ability to convert its asset base into revenue, profit, and returns on investment.

4. Results and Analysis

Descriptive Statistics of Variables

```
. sum
```

Variable	Obs	Mean	Std. dev.	Min	Max
ROS	213	.002242	.0021982	-.0008228	.0142917
LNSIZE	213	24.0394	1.395724	20.25312	27.5387
LNLIQ	213	.710974	.8233719	-.4107543	4.398157
LNLEV	213	-.7571754	.8154032	-4.398157	.1545786
TANG	213	.0358495	.0504298	0	.2335832
COGS	213	.8637047	.1602276	.0026224	1.211508
AUT	213	1.199244	.8614815	.0435044	3.564527

Multicollinearity Diagnostics

```
. vif
```

Variable	VIF	1/VIF
LNLEV	4.42	0.226057
LNLIQ	4.17	0.239994
COGS	1.46	0.682634
LNSIZE	1.38	0.726684
AUT	1.18	0.845315
TANG	1.13	0.882910
Mean VIF	2.29	

The Variance Inflation Factor (VIF) values for all independent variables ranged from approximately 1.13 to 4.42. The mean VIF was 2.29, which is well below the commonly accepted threshold of 5, indicating that multicollinearity is not a significant concern in the model.

Regression Results and Interpretation

```
. regress ROS LNSIZE LNLIQ LNLEV TANG COGS AUT
```

Source	SS	df	MS	Number of obs	=	213
				F(6, 206)	=	43.49
Model	.000572445	6	.000095408	Prob > F	=	0.0000
Residual	.00045197	206	2.1940e-06	R-squared	=	0.5588
				Adj R-squared	=	0.5460
Total	.001024415	212	4.8321e-06	Root MSE	=	.00148

ROS	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
LNSIZE	-.0000519	.0000855	-0.61	0.545	-.0002204	.0001167
LNLIQ	.0009802	.0002522	3.89	0.000	.000483	.0014774
LNLEV	-.0001746	.0002624	-0.67	0.507	-.0006919	.0003428
TANG	.0022538	.0021469	1.05	0.295	-.0019789	.0064864
COGS	-.0061047	.0007685	-7.94	0.000	-.0076198	-.0045897
AUT	.0002538	.0001284	1.98	0.049	6.16e-07	.0005071
_cons	.0075474	.0020735	3.64	0.000	.0034593	.0116354

None of the variables exceeded the VIF > 5 threshold, reinforcing the absence of serious multicollinearity. While LIQ (liquidity ratio) and LEV (financial leverage) may exhibit a slight degree of correlation, their VIF values remain within acceptable limits. Therefore, the regression analysis can proceed without the need to remove or adjust any of the independent variables.

With a sample of 213 observations, the regression model produced an R-squared value of 0.5588, indicating that approximately 55.88% of the variation in Return on Sales (ROS) is explained by the independent variables in the model. The adjusted R-squared is 0.5460, which accounts for the number of predictors used and confirms a consistent model fit. The F-statistic is F(6, 206) = 43.49 with a p-value < 0.0000, demonstrating the overall statistical significance of the model.

The intercept (_cons) is estimated at 0.00754, indicating the predicted value of ROS when all independent variables are equal to zero. The model is considered statistically sound and relatively effective in predicting financial performance (ROS).

Liquidity Ratio (LIQ): An increase of 1 unit in the previous year's liquidity ratio is associated with an increase of 0.00098 units in ROS in the following year.

Cost of Goods Sold (COGS): A decrease of 1 unit in the previous year's cost of goods sold ratio results in an increase of 0.0061 units in ROS.

Asset Utilization (AUT): An increase of 1 unit in asset utilization leads to an increase of 0.00253 units in ROS in the subsequent year.

Other variables, including firm size (SIZE), financial leverage (LEV), and asset tangibility (TANG), were found to have no statistically significant impact on financial performance in this model.

Based on the linear regression results, the study confirms that liquidity, cost of goods sold, and asset utilization from the previous fiscal year significantly affect financial decision-making and operational performance, and thus impact the financial performance and corporate income tax obligations in the subsequent year.

5. Application in Tax Forecasting

Indicator	Liquidity Ratio (LIQ)	Cost of Goods Sold (COGS)	Asset Utilization Ratio (AUT)	Corporate Income Tax Payable (VND)
2023	0,47733	0,90903	1,06497	129.753.799.572
2024	0,40951	0,91554	0,88842	378.568.987.945
Change 2023 - 2024	-14,21%	0,72%	-16,58%	

Using the model coefficients and financial dynamics from the 2023-2024 period, the regression model forecasts an estimated 0.7% increase in corporate income tax revenue in 2025.

The author applies the results of the multiple linear regression model to predict corporate income tax revenue for the sample of 213 enterprises in 2025. Key financial indicators were evaluated using median values, and the expected tax obligations were computed based on actual data from the enterprises' corporate income tax declarations.

Based on the Table Above

The author observes that in 2024, the liquidity ratio increased by 14.21%, the cost of goods sold (COGS) rose by 0.72%, while the asset utilization ratio declined by 16.58% compared to 2023. Based on these developments, the author applied the results of the multiple linear regression model and estimated a 0.71% increase in corporate income tax revenue in 2025. From the perspective of analyzing firm financial performance, it is projected that corporate income tax collections will rise 0.71% in 2025 relative to 2024.

6. Discussion

The results confirm the hypothesis that financial indicators can predict firm performance and indirectly estimate corporate tax obligations. Sectoral differences observed in the dataset suggest the need for industry-specific forecasting models.

Among the tested variables, LIQ (liquidity), COGS (cost of goods sold), and AUT (asset utilization) were found to have statistically significant effects on ROS (Return on Sales). In contrast, SIZE (firm size), LEV (financial leverage), and TANG (asset tangibility) did not show significant impacts within this model.

Model Limitations: This study has several limitations. It does not incorporate macroeconomic variables or account for policy changes, both of which can influence tax revenues. Future models should integrate a broader economic context and explore time-lagged effects to enhance forecasting precision.

7. Conclusion

This research proposes a practical and replicable quantitative model for forecasting corporate income tax revenues based on firm-level financial indicators. The regression analysis confirms that liquidity, cost of goods sold, and asset utilization have statistically significant impacts on financial performance (ROS), which is directly linked to tax obligations.

The multiple linear regression model demonstrated a relatively high explanatory power, highlighting its potential application in local-level fiscal planning. However, the model is constrained by a narrow sample scope and the absence of macroeconomic factors. Therefore, future studies should aim to develop industry-specific models and incorporate contextual and macro-level variables to improve accuracy and enhance policy relevance.

References:

Alarussi, A. S., & Alhaderi, S. M. (2018). Factors affecting profitability in Malaysia. Journal of Economic Studies, 45(3), 442-458.

Alghusin, N. A. S. (2015). The impact of financial leverage, growth, and size on profitability of Jordanian industrial listed companies. Research Journal of Finance and Accounting, 6(16), 86-93.

Chu, T. T. T., Nguyen, H. T., & Ngo, Q. T. (2015). Analysis of factors affecting financial performance: A case study in non-financial joint-stock companies listed on the Ho Chi Minh City stock exchange. Journal of Economics and Development, 215, 33-38.

Jenkins, G. P., & Kuo, C. Y. (2000). Tax Analysis and Revenue Forecasting. Harvard Institute for International Development.

Nguyen, D. T., Ta, H. T., & Nguyen, H. T. D. (2018). What determines the profitability of Vietnam commercial banks? International Business Research, 11(2), 231-245.

OECD. (2020). Corporate Tax Statistics. https://www.oecd.org/tax/tax-policy/corporate-tax-statistics/

Odusanya, I. A., Yinusa, O. G., & Ilo, B. M. (2018). Determinants of firm profitability in Nigeria: Evidence from dynamic panel models. SPOUDAI - Journal of Economics and Business, 68(1), 43-58.

Phu Hung Securities. (2019). 2018 Business Performance Review: Industry Textiles. Retrieved August 14, 2020, from https://www.phs.vn/data/research/PDF_Files/analysis_report/en/20190401/Textiles-20190330-E.pdf

Quan, N. M., & Ly, T. T. P. (2014). Analyze determinants of real estate firms' performance listed on the Vietnamese stock exchange. Can Tho University Journal of Science, 33, 65-71.

Wooldridge, J. M. (2012). Introductory Econometrics: A Modern Approach. Cengage Learning.