TAX ACCOUNTING
APPLIED JOURNAL
e-ISSN: 2986-0539
Vol. 5, Iss. 1, 2026, 01-14

An Evaluation of ARIMA, ETS, and HOLT-WINTERS Frameworks in Forecasting State Tax Revenue Collections in Indonesia

Ferry*, Qadri Fidienil Haq, Fajar Fathurrahman

Directorate General of Taxes, Ministry of Finance Indonesia

ABSTRACT

Introduction/Main Objectives: This study evaluates the performance of Auto-Regressive Integrated Moving Average (ARIMA), Error-Trend-Seasonality (ETS), and Holt-Winters frameworks in forecasting state tax revenue collections in Indonesia. **Background Problems:** Accurate tax revenue prediction is essential for prudent fiscal governance, yet comparative assessments of these methods remain limited in the existing literature. **Novelty:** this study contributes to the sparse public policy literature by testing the accuracy of three established frameworks within the context of Indonesian tax policy. **Research Methods:** We employed the period from January 2016 to March 2023 for model identification and the remaining twelve months for out-of-sample validation. **Finding/Results:** All three methods show good predictive accuracy for Indonesian tax revenues, with different strengths—ARIMA for PPh 21, ETS for PPN DN and as the overall best performer, and Holt-Winters for PPh OP. **Conclusion:** While these findings are specific to Indonesia, the methodological framework offers transferable insights for tax authorities seeking to enhance revenue forecasting and inform evidence-based policymaking.

ARTICLE INFO

Keywords:
forecasting, tax revenue, public policy, data analytics, public administration

Received: 31 Jan 2026
Revised: 4 Apr 2026
Accepted: 23 Apr 2026
Available Online: 10 July 2026
Published: 10 July 2026

JEL Code:
D13, I31, J22, K31

* Corresponding Author at Directorate General of Taxes, Ministry of Finance Indonesia
E-mail address: ferry81@kemenkeu.go.id (author#1), qadri.haq@kemenkeu.go.id (author#2), fajar.fathurrahman@kemenkeu.go.id (author#3).

1. Introduction

The study analyses tax revenue data in an attempt to inform technical analysis and develop sound public policy in Indonesia. Tax revenues are important for continuous development in Indonesia as they have continuously supported more than 70% of the public budget since 2009 (Central Bureau of Statistics Indonesia, 2025). Given these significant contributions, the planning of tax revenue collections is crucial to determine the direction of the public expenditures. Therefore, it is important to apply reliable forecasting so that the inspiration to improve public services can be secured without the urgency to issue additional government bonds (Vashishth et al., 2025).

Prior studies have attempted to analyze forecasting diverse tax types within a fiscal year. Ratshefola-Lerefolo & Mpeta (2023) and Tsoku et al (2024) evaluated the efficacy of Auto-Regressive-Integrated-Moving-Average (ARIMA), Bayesian Vector Auto-Regression (BVAR), and Error-Trend-Seasonality (ETS) in predicting South African state tax revenue. They suggested that ETS is optimal for total tax revenue estimation, while BVAR demonstrated superior accuracy for forecasting the majority of the main types of taxes (Yaqoob et al., 2023).

Furthermore, Ogbogu & Mancilla-David (2021) argued the one-parameter Double-Exponential-Smoothing (DES) model from Holt and DES from Brown as potential tools to forecast the total state tax revenue in Indonesia. This suggestion derived from each model's ability to achieve a low Mean-Absolute-Error (MAE) value. Furthermore, (Iraman2023) used the ARDL method to predict tax revenue using three variables: interest rates, exchange rates, and inflation. The results showed that macroeconomic factors were more influential in the short term, while time-series analysis remained the most appropriate for predicting certain types of taxes in the long term. On a more technical level, (Priandono & Eka Prasetya, 2025) identified a decomposition method to improve the attempt of tax revenue predictions.

The ARIMA model is a well-known method for forecasting revenue segmented by type or region as it has been applied in various countries, such in Middle Asia (Yadav et al., 2024), South Africa and Kenya (Kithure, 2021). Previous research has also applied ARIMA to predict tax revenues in Indonesia (Jayanti et al., 2025). Muslim (2022) attempted to predict import duty revenue and highlighted its effectiveness in this context. Further solidifying ARIMA's utility confirmed its accuracy in estimating VAT revenue in both Surabaya Karangpilang and Jakarta Setiabudi Dua Tax Offices (Fathoni & Saputra, 2023). Within the local tax domain, Pebriani Wahyu et al., 2022) explored alternative forecasting methods for Majalengka Regency. They compared Single Exponential Smoothing (SES), DES, and Holt-Winters. Their analysis identified Holt-Winters additive as the most effective approach to forecast the local tax revenue.

Given limited information available on examining the performance of ARIMA, ETS, and Holt-Winters in predicting state tax revenue collections in Indonesia, the study contributes to the relatively limited literature on public policy and aims to answer the following research questions: what is the accuracy level of the three frameworks in predicting various tax revenues in Indonesia during the period from April 2023 to March 2024, and which framework generates the most accurate predictions among the three in the testing period.

2. Literature Review

ARIMA framework combines the analysis of past values (auto-regressive), difference order (integrated), and past residuals (moving-average) to determine the autocorrelation function within the time series (Hyndman, 2021). The framework is written as ARIMA (p, d, q)(P, D, Q)m, in which p is an order of auto-regressive model, d is a degree of differencing, and q is a moving-average component (Asimopolos et al., 2020; Jiang et al., 2025; Verma et al., 2024). A prerequisite to perform ARIMA frameworks is the stationary of the data. Hence, differencing the non-stationary data to some degree (d) is a must. The (P, D, Q) is used for the seasonal part. For instance, ARIMA $(1,1,1)(1,1,1)_{12}$ means the model applied one period time lag of the AR model, one step differencing, and one period time lag of the MA model. The number of 12 explains the characteristic of monthly data.

The next framework, ETS, is based on exponential smoothing methods(Pedregal et al., 2025). It includes errors, trends, and seasonality parts to determine optimal performance models to simulate the data (Qi et al., 2023). Based on the three main parameters, ETS can be classified into three criteria: Additive (A); Multiplicative (M); and None (N). To choose the fitted model, the residual order should form white noise order or display a constant variance and be distributed independently across the time series. ETS has been considered appropriate to handle any error, which may appear in either additive error or multiplicative error (Svetunkov & Boylan, 2024). ETS with additive error feature sums up its three equations (level, trend, and seasonality) while ETS with multiplicative feature multiplies its three equations. ETS Model with none criteria in trend or seasonal do not include the related equation. For example, trend equations and seasonal equations are not included in ETS (A, N, N).

The third model, Holt-Winters, comprises one prediction equation and three smoothing equations (smoothing level (denoted as α), smoothing trend (β), and smoothing seasonality (γ)) to reduce random variances while adjusting seasonal trends (Liu & Wu, 2022; Nan et al., 2025; Wongoutong, 2021a). It is also called triple exponential smoothing because of the capability to identify the three components of time series (level, trend, and seasonality) simultaneously (Anthonysamy et al., 2023). Compared to ETS, Holt-Winters framework does not include the error in the equation. Holt-Winters selects a variety of weights of data from each season and sensibly forecasts future emerging trends (Wongoutong, 2021b). Both the level and the slope of the trends are in the range of 0 to 1, where a value nearer to zero indicates that the predictions are based on latest observations. The framework of Holt-Winters has two features: additive and multiplicative (Wongoutong, 2021b, 2021a). The former is applied in constant seasonal data series whereas the latter in the case of changes in the seasonal change move proportionally to the level of data series

3. Method, Data, and Analysis

The authors collected monthly tax revenue data during the period from January 2016 to March 2024, which was retrieved from the data warehouse of the Directorate General of Taxes Indonesia on 8 May 2024. The absolute values of each tax revenue were then transformed into log values to enhance the stability of the datasets by decreasing the influence of extreme values in each time series (Hassani et al., 2020).

Further, the dataset was divided into two sets: the first dataset, from January 2016 to March 2023, was used as a training set, in which various frameworks, namely ARIMA, ETS, and Holt-Winters were applied to identify the optimal model to describe the dataset. The following 12-month period dataset, from April 2023 to March 2024, was used as an evaluation set.

Following the design of the Indonesian state budget revenue, the study classifies the type of tax revenue into five categories (non-oil income taxes, value-added taxes (VAT), land and building taxes, other taxes, and oil taxes), and further extends it into 17 types, which consist as follows: taxes on employment (PPh 21); taxes on specific trade transactions (PPh 22); taxes on import activities (PPh 22 Impor); taxes on certain incomes (PPh 23); taxes installments of individuals (PPh 25 OP); taxes installments of entities (PPh 25 Badan); taxes withheld from Indonesian-sourced incomes (PPh 26); taxes with the final settlements (PPh Final); other non-oil taxes (PPh Nonmigas Lainnya); VAT on Domestic Products (PPN DN); VAT on Import Products (PPN Impor); sales taxes on luxury items domestic (PPnBM DN); sales taxes on luxury items import (PPnBM Impor); other VAT and sales taxes on luxury items (PPN PPnBM Lainnya); Land and Building Taxes (PBB); other taxes (Pajak Lainnya); oil taxes (PPh Migas).

To complement the analysis, we included the sum of the taxes into the total tax so that there are 18 sets of time series. It is also necessary to note that the government has launched a series of tax programs such as the tax amnesty program in 2016 and the voluntary disclosure program in 2022 to enhance taxpayers' compliance. Both programs successfully accumulated additional tax revenues of IDR 135 trillion and IDR 61 trillion in the respective years (Directorate General of Taxes Indonesia, 2018;

2023). Accordingly, for the purpose of the analysis and given that such programs are not repeated in the following years, the amounts are excluded from the dataset in the respective years.

To run the evaluations, we used Python programming language in Python 3.12.2 interpreter, more specifically due to Python's capability in analysing and selecting the most optimal model which has the least Akaike Information Criterion (AIC), an indicator error, in an automated fashion. Previous study such as: pandas 2.2.0, numpy 1.24.3, matplotlib 3.8.2, seaborn 0.13.0, pmdarima 2.0.4, sklearn 1.4.0, and statsmodels 0.14.1 were added to perform the analyses, and evaluate the performances (Rahal et al., 2022). To ensure replicability, we include the applied analysis codes in Appendix. Following the prediction analyses, we adopted three established measurements, namely MAE, Mean-Squared-Error (MSE), and Root-Mean-Squared-Error (RMSE), to assess the accuracy of the predictions from the three frameworks. We then selected the optimal framework that recorded the least indicator errors. The formula of the evaluations is as follows:

$$\text{MAE} = \sum (|\text{test} - \text{prediction}|) \times (1/\text{test size}) \quad (1)$$

$$\text{MSE} = \sum (|\text{test} - \text{prediction}|)^2 \times (1/\text{test size}) \quad (2)$$

$$\text{RMSE} = \sqrt{[\sum (|\text{test} - \text{prediction}|)^2 \times (1/\text{test size})]} \quad (3)$$

To sum up, the analysis in the study is as follows: (1) plot the observed time series, including the trend, seasonality, and residual components; (2) check the stationarity by running the unit root tests, namely Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test; (3) select the optimal model with the least AIC in the training dataset; and (4) run the selected model in the testing dataset and evaluate the accuracy based on the MAE, MSE, and RMSE

4. Result and Discussion

Table 1 displays the descriptive statistics of tax revenues (log values) during the period from January 2016 to March 2024. It includes minimums, averages, standard deviations, and maximums. It can be seen in the minimum column that the tax which recorded the lowest collection during the period was PPN_PPnBM_Lain at 9.02 whereas the highest tax collected in the similar period was PPh_Badan at 14.13 (in the maximum column). Furthermore, PPN_PPnBM_Lain and PBB were the taxes which displayed high volatility attributes as the standard deviations were 0.88 and 0.70, respectively.

To understand more about the features of the data, we depicted the data in Figure 1. We detailed the visual into four parts: the observed, trend, seasonality, and residual. For simplicity and illustrative purposes, we compared and plotted two decomposition models: additive and multiplicative models for the total tax only (see Figure 1). Each of the figures in Figure 1 consists of observed (top), trend (second top), seasonality (third top), and residuals (bottom).

Table 1. Descriptive Statistics

Tax	mean	std	min	q-1	q-2	q-3	max
PPh_21	13.08	0.13	12.81	12.99	13.07	13.17	13.45
PPh_22	12.23	0.20	11.77	12.08	12.21	12.43	12.70
PPh_22_Impor	12.59	0.21	11.97	12.52	12.64	12.74	12.85
PPh_23	12.54	0.11	12.30	12.47	12.52	12.63	12.85
PPh_OP	11.74	0.35	11.33	11.59	11.66	11.75	12.85
PPh_Badan	13.31	0.22	12.88	13.18	13.28	13.39	14.13
PPh_26	12.67	0.21	12.18	12.50	12.70	12.82	13.14
PPh_Final	12.98	0.09	12.83	12.92	12.95	13.01	13.38
PPh_Nonmigas_Lain	9.95	0.22	9.41	9.81	9.91	10.05	10.64
PPN_DN	13.58	0.14	13.31	13.48	13.54	13.69	13.91
PPN_Impor	13.18	0.13	12.92	13.07	13.18	13.29	13.42
PPnBM_DN	11.91	0.31	10.67	11.74	11.99	12.09	12.73
PPnBM_Impor	11.53	0.21	10.78	11.43	11.57	11.65	12.00
PPN_PPnBM_Lain	10.68	0.88	9.02	9.93	10.44	11.59	11.86
PBB	11.72	0.70	10.51	11.10	11.63	12.25	13.16
PPh_Migas	12.62	0.19	12.02	12.52	12.63	12.73	13.16
Pajak_Lain	11.85	0.15	11.58	11.74	11.83	11.93	12.40
Total	14.08	0.12	13.86	13.98	14.06	14.19	14.44

Source: Adapted From Data Analysis Results.

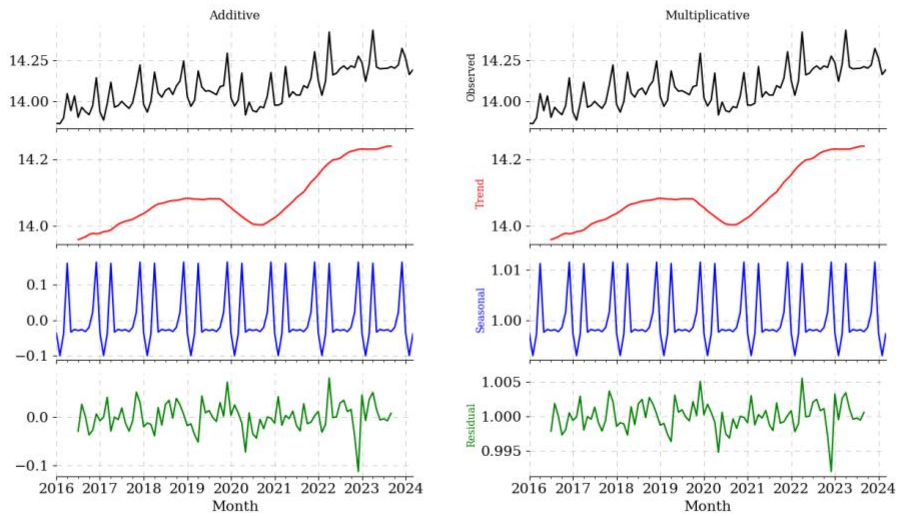

Figure 1. Seasonal Decomposition of Total Tax: Additive and Multiplicative Models
Source: Adapted From Data Analysis Results

Three observations related to the trend, seasonality, and residual can be noted in Figure 1. First, it shows a visible downward trend during the period 2020−2021 when the Covid-19 pandemic started to spread in Indonesia before going up in the following year. Secondly, there is an obvious seasonality attribute in the total tax data as the seasonality line plot keeps repeating across the time series. Finally, there is a difference between the two plots in the residual component: the residuals in the additive model (left panel) scatter in a distributed manner across the data set whereas those in the multiplicative model (right panel) are set constantly at around 1.0. In line with this result, we thus run exponential smoothing analyses (ETS and Holt-Winters) with additive features.

After determining the compositional features and confirming seasonality, we progressed to the second step, which is running ADF and KPSS tests. This is important to determine further analyses, because when a dataset is already stationary, a simple analysis (for example: average growth analysis) is sufficient to run the predictions. Conversely, when a dataset exhibits non-stationary features, it is necessary to address the non-stationary aspect before producing any prediction analysis. Hence, ARIMA framework, which has been considered appropriate to address the non-stationary issue by differencing, was included in the analysis.

The results of ADF and KPSS tests are summarised in Table 2. For example, in total tax, the outcome of ADF statistic is −0.61 and p-value of 0.87. The results indicate that we can reject the hypothesis that the dataset is stationary. A similar interpretation result can be seen in the KPSS test for total tax, in which the KPSS statistic showed 1.29 and p-value of 0.01: in contrast with the ADF test, the significant of the p-value indicates the non-stationarity feature of the dataset. Hence, both tests confirm that the dataset of total tax is non-stationary. Similar interpretations were applied to the remainder of the taxes.

Table 2. Unit Root Test Results – ADF and KPSS

Tax	ADF Stat	ADF p-value	KPSS Stat	KPSS p-value
PPh_21	1.104	0.995	1.667	0.01
PPh_22	−0.986	0.758	1.529	0.01
PPh_22_Impor	−2.527	0.109	0.306	0.1
PPh_23	−0.468	0.898	1.466	0.01
PPh_OP	−2.57	0.099	0.885	0.01
PPh_Badan	−1.5	0.534	0.8	0.01
PPh_26	−1.135	0.701	1.392	0.01
PPh_Final	−0.519	0.888	0.277	0.1
PPh_Nonmigas_Lain	−0.34	0.92	0.876	0.01
PPN_DN	0.267	0.976	1.392	0.01
PPN_Impor	−1.315	0.622	1.221	0.01
PPnBM_DN	−2.283	0.178	0.223	0.1
PPnBM_Impor	−0.646	0.86	0.434	0.063
PPN_PPnBM_Lain	−0.711	0.844	1.253	0.01
PBB	−0.885	0.793	0.201	0.1
PPh_Migas	−2.12	0.236	0.43	0.064
Pajak_Lain	−2.704	0.073	1.282	0.01
Total	−0.611	0.868	1.29	0.01

Source: Adapted From Data Analysis Results

The following two steps are selecting the optimal models and running each of the optimal models to get a set of predictions. Utilizing Python's capacity, the analyses were undertaken simultaneously

(Altenkort et al., 2024; Nava & Lee, 2023). The summary of the analysis results is listed in Table 3 and the frameworks' reviews by the reference of MAE, MSE, and RMSE are outlined in Table 4. The accuracy results are then discussed in turn.

Table 3. Training Dataset Results

Tax	ARIMA (p,d,q)(P,D,Q) : AIC	ETS α ; β ; γ : AIC	HW α ; β ; γ : AIC
PPh_21	(0,1,1)(1,1,0)[12]: −210	0.06; 0.05; 0.00: −242	0.14; 0.00; 0.42: −489
PPh_22	(0,1,1)(0,1,1)[12]: −130	0.25; 0.02; 0.00: −163	0.25; 0.00; 0.00: −418
PPh_22_Impor	(0,1,0)(2,1,0)[12]: −84	0.99; 0.00; 0.00: −112	0.97; 0.00; 0.00: −365
PPh_23	(0,1,1)(0,1,1)[12]: −193	0.05; 0.04; 0.00: −239	0.17; 0.00; 0.00: −492
PPh_OP	(0,1,1)(2,1,0)[12]intercept: −43	0.00; 0.00; 0.00: −82	0.08; 0.08; 0.00: −327
PPh_Badan	(1,1,0)(0,1,0)[12]: −138	0.46; 0.00; 0.54: −142	0.46; 0.00; 0.54: −396
PPh_26	(2,1,1)(2,1,0)[12]: −55	0.00; 0.00; 0.00: −87	0.00; 0.00; 0.00: −341
PPh_Final	(0,1,1)(0,1,1)[12]: −156	0.00; 0.00; 0.00: −195	0.00; 0.00; 0.00: −446
PPh_Nonmigas_Lain	(3,1,0)(2,1,0)[12]: −21	0.08; 0.00; 0.00: −58	0.00; 0.00; 0.00: −318
PPN_DN	(0,1,1)(0,1,1)[12]: −256	0.47; 0.03; 0.00: −308	0.44; 0.00; 0.00: −564
PPN_Impor	(3,1,1)(0,1,0)[12]: −230	0.25; 0.25; 0.00: −259	0.47; 0.00; 0.00: −506
PPnBM_DN	(0,1,1)(0,1,1)[12]: 26	0.47; 0.00; 0.00: 19	0.47; 0.00; 0.00: −233
PPnBM_Impor	(0,1,1)(0,1,1)[12]: −66	0.36; 0.00; 0.00: −88	0.37; 0.00; 0.00: −339
PPN_PPnBM_Lain	(0,1,1)(2,1,0)[12]: 101	0.62; 0.00; 0.00: 95	0.61; 0.00; 0.00: −158
PBB	(2,1,0)(2,1,0)[12]: 132	0.00; 0.00; 0.00: 121	0.00; 0.00; 0.00: −132
PPh_Migas	(1,1,0)(2,1,0)[12]: −108	0.54; 0.00; 0.00: −128	0.55; 0.00; 0.00: −380
Pajak_Lain	(0,1,1)(2,1,0)[12]: −61	0.13; 0.00; 0.00: −84	0.00; 0.00; 0.00: −341
Total	(0,1,1)(0,1,1)[12]: −234	0.35; 0.08; 0.00: −282	0.43; 0.00; 0.00: −535

Source: Adapted From Data Analysis Results

Table 4 reveals that the analysis results by the reference of MAE correspond to the results by the reference of MSE and RMSE consecutively, which means a framework which generates the least error in MAE also produces the least errors in MSE and RMSE proportionately. Hence, we carried on the analysis discussions by the reference of MAE only. The first observation in Table 4 finds that ETS generated the least error in the analysis of PPN_Impor (1.97%) while ARIMA and Holt-Winters showed the most accurate prediction in the analysis of PPh_23 (2.71% and 2.78%, respectively).

In terms of the average error among the three frameworks: PPh_23 and PPN_DN were the taxes which proportionately produced the least errors in all three frameworks (on average 2.76% and 3.32%, respectively) whereas PBB and PPN_PPnBM_Lain were the type of taxes which equally generated the biggest errors in the three frameworks (the averages of MAE are 32.26% and 23.02%, respectively). Perhaps, the poor prediction results due to the high volatility features of PBB and PPN_PPnBM_Lain (see Table 1).

Table 4. Evaluation Results by the references of MAE, MSE, and RMSE

Tax	ARIMA	ETS	HW
PPh_21 MAE	3.11%	5.90%	4.44%
PPh_21 MSE	0.18%	0.49%	0.33%
PPh_21 RMSE	4.26%	7.00%	5.76%
PPh_22 MAE	8.28%	5.54%	7.79%
PPh_22 MSE	0.89%	0.50%	0.88%
PPh_22 RMSE	9.44%	7.04%	9.36%
PPh_22_Impor MAE	7.30%	5.50%	7.92%
PPh_22_Impor MSE	0.88%	0.36%	0.70%
PPh_22_Impor RMSE	9.40%	6.01%	8.36%
PPh_23 MAE	2.71%	2.79%	2.78%
PPh_23 MSE	0.14%	0.13%	0.14%
PPh_23 RMSE	3.73%	3.65%	3.74%
PPh_OP MAE	9.49%	8.59%	7.99%
PPh_OP MSE	2.24%	1.36%	1.36%
PPh_OP RMSE	14.96%	11.66%	11.64%
PPh_Badan MAE	21.23%	10.44%	11.73%
PPh_Badan MSE	5.79%	1.62%	1.81%
PPh_Badan RMSE	24.06%	12.71%	13.45%
PPh_26 MAE	9.80%	12.43%	11.85%
PPh_26 MSE	1.28%	2.25%	2.00%
PPh_26 RMSE	11.32%	15.00%	14.13%
PPh_Final MAE	6.58%	6.94%	6.80%
PPh_Final MSE	0.51%	0.61%	0.59%
PPh_Final RMSE	7.17%	7.81%	7.69%
PPh_Nonmigas_Lain MAE	14.58%	13.25%	9.99%
PPh_Nonmigas_Lain MSE	2.65%	2.08%	1.40%
PPh_Nonmigas_Lain RMSE	16.26%	14.42%	11.82%
PPN_DN MAE	3.80%	2.83%	3.35%
PPN_DN MSE	0.20%	0.14%	0.19%
PPN_DN RMSE	4.49%	3.72%	4.36%
PPN_Impor MAE	5.16%	1.97%	5.70%
PPN_Impor MSE	0.40%	0.10%	0.41%
PPN_Impor RMSE	6.29%	3.08%	6.42%
PPnBM_DN MAE	15.73%	14.89%	15.69%
PPnBM_DN MSE	14.84%	3.11%	6.66%
PPnBM_DN RMSE	38.52%	17.63%	25.81%
PPnBM_Impor MAE	8.70%	11.76%	12.16%
PPnBM_Impor MSE	4.03%	3.55%	4.11%
PPnBM_Impor RMSE	20.06%	18.85%	20.28%
PPN_PPnBM_Lain MAE	34.52%	13.90%	20.62%
PPN_PPnBM _Lain MSE	1.62%	2.82%	2.91%
PPN_PPnBM_Lain RMSE	12.73%	16.79%	17.06%
PBB MAE	32.92%	33.13%	30.74%
PBB MSE	16.89%	17.09%	14.83%
PBB RMSE	41.09%	41.33%	38.50%
PPh_Migas MAE	11.58%	8.09%	10.92%
PPh_Migas MSE	1.82%	0.90%	1.58%
PPh_Migas RMSE	13.49%	9.49%	12.59%
Pajak_Lain MAE	9.70%	6.97%	7.57%
Pajak_Lain MSE	2.28%	1.60%	1.64%
Pajak_Lain RMSE	15.10%	12.67%	12.82%
Total MAE	5.51%	3.02%	4.32%
Total MSE	0.36%	0.15%	0.26%
Total RMSE	6.04%	3.93%	5.09%

Source: Adapted From Data Analysis Results

Further, the evaluations of MAE indicate that ARIMA outperforms the two frameworks in the predictions of PPh_21 (3.11%), PPh_23 (2.71%), PPh_26 (9.80%), PPh_Final (6.58%), and

PPnBM_Impor (8.70%). A further examination discovers that Holt-Winters works better compared the other two in predicting PPh_OP (7.99%), PPh_Nonmigas_Lain (9.99%), and PBB (30.74%). Finally, ETS generates less error rates in comparison of ARIMA and Holt-Winters frameworks in analysing PPh_22 (5.54%), PPh_22_Impor (5.50%), PPh_Badan (10.44%), PPN_DN (2.83%), PPN_Impor (1.97%), PPnBM_DN (14.89%), PPN_PPnBM_Lain (13.90%), PPh_Migas (8.09%), Pajak_Lain (6.97%), and Total (3.02%). Thus, it can be concluded that among the three frameworks in 18 time series of various taxes in the study: ETS is the best in 10 analyses; ARIMA comes in second in 5 predictions; and Holt-Winters gets the best results in 3 estimates.

Compared to ARIMA and Holt-Winters, ETS performs very well in predicting the taxes of PPh_Badan, PPN_DN, and PPN_Impor (Lima et al., 2024). Given that these taxes have been substantially contributing to the total tax revenues in Indonesia, ETS is superior in predicting the total tax (Rosenblad, 2021). Similarly, ARIMA generates excellent results in predicting PPh_21 and PPh_23. Nevertheless, ARIMA performs worst in 10 of 18 estimates, more particularly in predicting PPh_Badan and total tax. Finally, despite being superior in predicting three taxes only, Holt-Winters provides such comparable results with those of ARIMA and ETS that it performs worst only in PPh_22_Impor, PPN_Impor, and PPnBM_Impor.

To provide the complete evaluation in visuals, Figure 2 displays the evaluation plot of ARIMA vs ETS vs HW in PPh_Nonmigas & PPh_Migas and Figure 3 illustrates the evaluation plot of ARIMA vs ETS vs HW in PPN & Pajak_Lain. Overall, each of the three frameworks has its own advantage in predicting tax revenue collections in Indonesia. For example: ARIMA works best in predicting PPh_21 whereas ETS performs superiorly in predicting PPN_DN and Holt-Winters provides excellent predictions in PPh_OP. Hence, a strategic way to obtain the optimal predictions can be achieved by combining the frameworks of ARIMA, ETS, Holt-Winters.

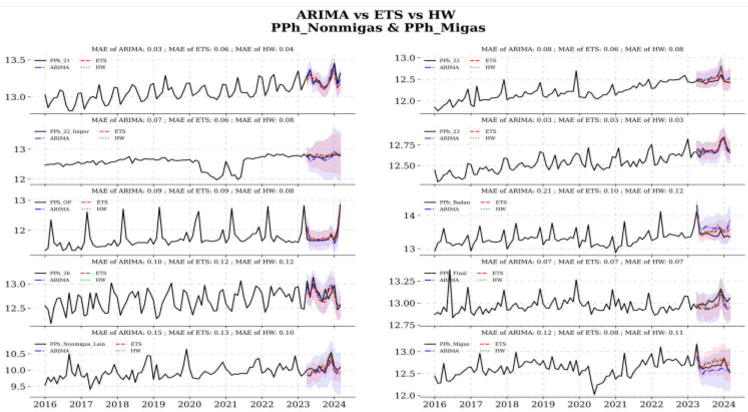

Figure 2. Evaluation Plot of ARIMA vs ETS vs HW: PPh_Nonmigas & PPh_Migas
Source: Adapted From Data Analysis Results

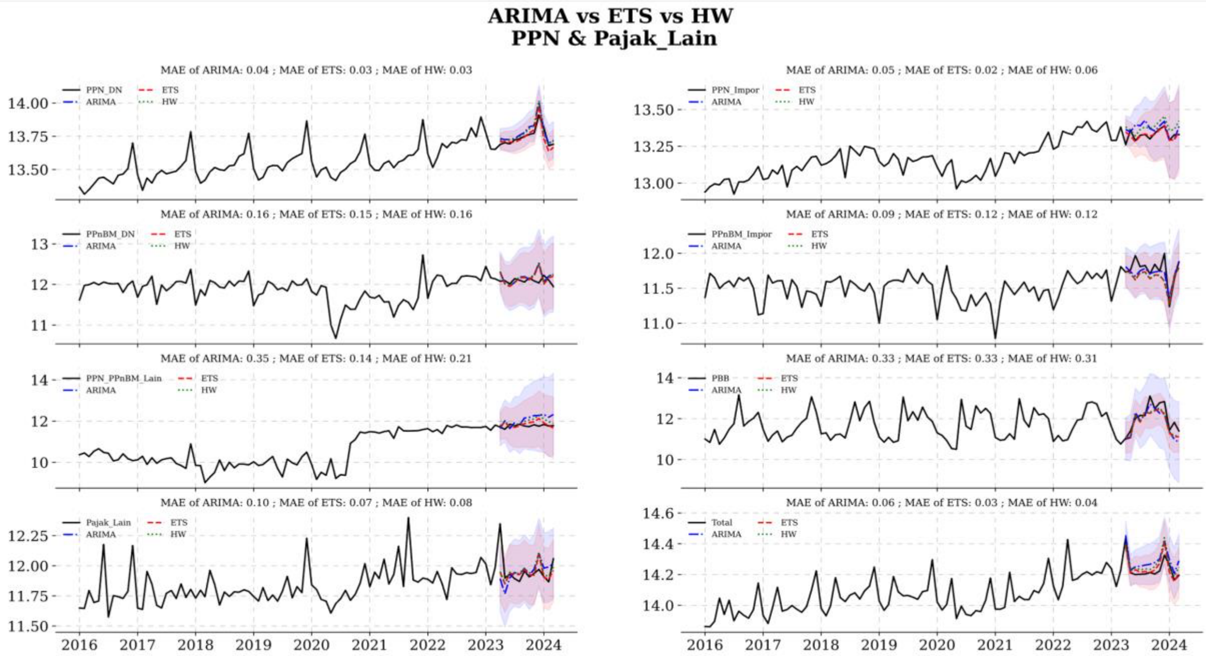

Figure 3. Evaluation Plot of ARIMA vs ETS vs HW: PPN & Pajak_Lain
Source: Adapted From Data Analysis Results

5. Conclusion and Suggestion

We have analysed the accuracy of ARIMA, ETS, and Holt-Winters in predicting Indonesian tax revenue collections during the period from April 2023 to March 2024. The evaluation revealed the optimal least average error of predictions produced in the analysis of PPN_Impor (1.97%) whereas the highest average deviations generated in the prediction of PBB (30.74%). We also observed that ETS is more superior than the other two frameworks in ten of 18 estimates. Given the comparative advantage of each framework in a certain analysis, the combination of the three frameworks can generate more optimal tax predictions

The findings of the study confirm that ARIMA, ETS, Holt-Winters generate sensible accuracies in predicting state tax revenues in Indonesia. The study also indicates that ETS performs best among the three by recording the least indicator errors in majority estimates. The results can be used to provide a range of statistical analyses which in turn allow public policy enhancements.

The study, despite the contributions, is not free from theoretical as well as practical limitations, specifically related to the Covid−19 pandemic situation in Indonesia. Despite the observed influence of the pandemic on various economic sectors in Indonesia, we could not reach a firm consensus about the clear-cut period of Covid−19 pandemic in the study due to the relatively limited literature support and the intrinsic features of tax revenue collections in Indonesia. For example, in the case of taxes installments of individuals and entities (PPh_OP and PPh_Badan), these taxes are due and submitted based on the calculation of taxable incomes from the previous year. As a result, the prediction results should be interpreted with caution, especially when evaluating these installment taxes.

Future studies could apply the technical framework used in this study and replicate the analysis to forecast other macroeconomic indicators related to public policy settings, in various fields, and across different disciplines. Hence, the technical analysis applied in this study can be tested on other state revenue sources, such as customs and excise. Since such revenues are also related to the public economy, the analysis will improve the public policy regulations.

Despite the limitations and recommended areas for further studies, this study has contributed additional insight into the underexplored analyses of state tax revenue collections in Indonesia, by identifying the application frameworks that assist in informing public policy administrations.

The findings of the study confirm that ARIMA, ETS, Holt-Winters generate sensible accuracies in predicting state tax revenues in Indonesia. The study also indicates that ETS performs best among the three by recording the least indicator errors in majority estimates. The results can be used to provide a range of statistical analyses which in turn allow public policy enhancements.

Reference

Altenkort, L., Clarke, D. A., Goswami, J., & Sandmeyer, H. (2024). Streamlined data analysis in Python. *Proceedings of Science*, *453*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209995097&partnerID=40&md5=afa5a1312637bef02aa67515e417b15b

Anthonysamy, V., Babu, K., & Chesneau, C. (2023). Heuristics approach for the Holt-Winters multiplicative method with new initial values. *Model Assisted Statistics and Applications*, *18*(2), 165–170. https://doi.org/10.3233/MAS-221415

Asimopolos, L., Asimopolos, N.-S., & Asimopolos, A.-A. (2020). *STUDY ABOUT AUTO-REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL USED FOR FORECAST OF STRONG GEOMAGNETIC DISTURBANCES*. 723–730. https://doi.org/10.5593/sgem2020/6.1/s28.094

Central Bureau of Statistics Indonesia. (2025). *Actual Government Revenues (Billion Rupiahs), 2007–2024*. Https://Www.Bps.Go.Id.

Fathoni, M. I., & Saputra, A. (2023). Forecasting Value-Added Tax (VAT) revenue using Autoregressive Integrated Moving Average (ARIMA) Box-Jenkins method. *Scientax*, *4*(2), 205–218. https://doi.org/10.52869/st.v4i2.568

Hassani, H., Yeganegi, M. R., Khan, A., & Silva, E. S. (2020). The Effect of Data Transformation on Singular Spectrum Analysis for Forecasting. *Signals*, *1*(1), 4–25. https://doi.org/10.3390/signals1010002

Hyndman, R. J. , & A. G. (2021). *Forecasting: principles and practice* (Third).

Iraman, E. (2023). Tax revenue forecasting: an empirical evidence using autoregressive-distributed lag (ARDL). *Qeios. Doi*, *10*, B21WPB.

Jayanti, D., Sulistyo, S., & Santosa, P. I. (2025). Forecasting Tax Revenue in Indonesia using Single and Hybrid Methods. *2025 International Conference on Advancement in Data Science, E-Learning and Information System (ICADEIS)*, 1–6. https://doi.org/10.1109/ICADEIS65852.2025.10933293

Jiang, W., Tan, Y., Du, Y., Chang, L., Chen, C., & Dong, W. (2025). The Time Series Prediction of Tourist Populations Based on ARIMA Model in Xizang. *Proceedings of the 2025 International Conference on Big Data, Artificial Intelligence and Digital Economy*, 188–195. https://doi.org/10.1145/3767052.3767081

Kithure, M. (2021). Use of Time Series Models on Forecasting of Value Added Tax Revenue in Kenya Revenue Authority. *African Tax and Customs Review*, *4*(2). https://atcr.kra.go.ke/index.php/atcr/article/view/81

Lima, S., Gonçalves, A. M., & Costa, M. (2024). Predictive accuracy of time series models applied to economic data: the European countries retail trade. *Journal of Applied Statistics*, *51*(9), 1818–1841. https://doi.org/10.1080/02664763.2023.2238249

Liu, L., & Wu, L. (2022). Holt–Winters model with grey generating operator and its application. *Communications in Statistics - Theory and Methods*, *51*(11), 3542–3555. https://doi.org/10.1080/03610926.2020.1797804

Muslim, H. A. (2022). Time Series Analysis for Customs Revenue Prediction using Arima Model in Python. *Journal of Information Technology and Its Utilization*, *5*(2), 47–55. https://doi.org/10.56873/jitu.5.2.4927

Nan, T., Liu, Z., Wu, X., Jiao, Z., & Sun, D. (2025). *Electricity Material Demand Forecasting Based on Holt-Winters and Random Forest*. https://doi.org/10.3233/ATDE250821

Nava, J., & Lee, H. (2023). mpiPython: Prospects for Node Performance. *2023 6th International Conference on Information and Computer Technologies (ICICT)*, 182–187. https://doi.org/10.1109/ICICT58900.2023.00038

Ogbogu, G., & Mancilla-David, F. (2021). Forecast-Aided State Estimation Smoothing Parameter Optimization using Winter-Holt's Methodology and RMSE. *2020 52nd North American Power Symposium (NAPS)*, 1–6. https://doi.org/10.1109/NAPS50074.2021.9449701

Pebriani Wahyu, F., Dwi Rahmawati, I., & Umam, K. (2022). Identifying Best Method for Forecasting Tax Income using Time Series Analysis. *Iapa Proceedings Conference*, 60. https://doi.org/10.30589/proceedings.2022.683

Pedregal, D. J., Trapero, J. R., & Holgado, E. (2025). Using Tobit exponential smoothing to enhance demand planning in the presence of censored data. *International Journal of Production Research*, 1–13. https://doi.org/10.1080/00207543.2025.2600521

Priandono, Y. A., & Eka Prasetya, M. (2025). Dekomposisi Penerimaan Pajak di Indonesia untuk Meningkatkan Peramalan Estimasi Basis Pajak. *Owner*, *9*(1), 130–139. https://doi.org/10.33395/owner.v9i1.2388

Qi, L., Li, X., Wang, Q., & Jia, S. (2023). fETSmcs: Feature-based ETS model component selection. *International Journal of Forecasting*, *39*(3), 1303–1317. https://doi.org/10.1016/j.ijforecast.2022.06.004

Rahal, R.-M., Hamann, H., Brohmer, H., & Pethig, F. (2022). Sharing the Recipe: Reproducibility and Replicability in Research Across Disciplines. *Research Ideas and Outcomes*, *8*. https://doi.org/10.3897/rio.8.e89980

Ratshefola-Lerefolo, L., & Mpeta, K. N. (2023). Analysis and Forecasting of Final Household Expenditure in South Africa Using Box-Jenkins – ARIMA Model. *5th International Conference on Intelligent Computing and Optimization, ICO 2022*, 1192–1196. https://doi.org/10.1007/978-3-031-19958-5_111

Rosenblad, A. K. (2021). Accuracy of automatic forecasting methods for univariate time series data: A case study predicting the results of the 2018 Swedish general election using decades-long data series. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *7*(3), 475–493. https://doi.org/10.1080/23737484.2021.1964407

Svetunkov, I., & Boylan, J. E. (2024). Staying positive: challenges and solutions in using pure multiplicative ETS models. *IMA Journal of Management Mathematics*, *35*(3), 403–425. https://doi.org/10.1093/imaman/dpad028

Tsoku, J. T., Metsileng, D., & Botlhoko, T. (2024). A Hybrid of Box-Jenkins ARIMA Model and Neural Networks for Forecasting South African Crude Oil Prices. *International Journal of Financial Studies*, *12*(4), 118. https://doi.org/10.3390/ijfs12040118

Vashishth, T. K., Sharma, V., Sharma, K. K., Ahamad, S., & Kaushik, V. (2025). Financial Forecasting with Convolutional Neural Networks (CNNs): Trends and Challenges. In *Shaping Cutting-Edge Technologies and Applications for Digital Banking and Financial Services* (pp. 62–81). Productivity Press. https://doi.org/10.4324/9781003501947-4

Verma, R., Pandey, S., & Rup, S. (2024). Traffic Flow Prediction Using LSTM, ARIMA & SARIMA Model for Intelligent Transportation System. *2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT)*, 1–6. https://doi.org/10.1109/IC-SIT63503.2024.10862828

Wongoutong, C. (2021a). Improvement of the holt-winters multiplicative method with a new initial value settings method. *Thailand Statistician*, *19*(2), 280–293. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85104045713&partnerID=40&md5=e18f5d8ccade2f3309ad534ec32a244d

Wongoutong, C. (2021b). The effect on forecasting accuracy of the holt-winters method when using the incorrect model on a non-stationary time series. *Thailand Statistician*, *19*(3), 565–582. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85111273963&partnerID=40&md5=5e60770c0c23aab7ffa0ba78cc98919b

Yadav, D. K., K, S., & Goswami, L. (2024). Retracted: Autoregressive Integrated Moving Average Model for Time Series Analysis. *2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC)*, 1–6. https://doi.org/10.1109/ICOCWC60930.2024.10470488

Yaqoob, I. A. M., Kasmiran, K. A., Aris, T. N. M., Husin, N. A., & Sharum, M. Y. (2023). IMPROVING CASH FORECASTING PERFORMANCE USING A NEW HYBRID-ANN-ARIMA-EM MODEL. *Journal of Theoretical and Applied Information Technology*, *101*(18), 7192–7204. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85175554643&partnerID=40&md5=c2785c3673f3c270cddeddfd74fc62b3