Forecasting Taxpayer Registration Using ARIMA Models: A Case Study of KPP Pratama Meulaboh

Riska Mulyani(1), Wanda Surianto(2), M. Riswan(3), Irham Akbar(4), Noor Aisah(5)

[1,2,4,5]Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Syiah Kuala, Banda Aceh, Indonesia

[3]Departemen Ilmu Kelautan, Fakultas Kelautan dan Perikanan, Universitas Syiah Kuala, Banda Aceh, Indonesia

Jl. Teuku Nyak Arief No.441, Kopelma Darussalam, Kec. Syiah Kuala, Kota Banda Aceh, Aceh 23111

e-mail: riskamulyani@usk.ac.id(1), wanda.surianto@usk.ac.id(2), m.riswan@usk.ac.id(3) irham23@mhs.usk.ac.id(4), noor23@mhs.usk.ac.id(5)

ABSTRAK

Jumlah wajib pajak (WP) terdaftar merupakan indikator penting dalam menilai efektivitas administrasi perpajakan serta tingkat partisipasi masyarakat dalam memenuhi kewajiban perpajakan. KPP Pratama Meulaboh sebagai salah satu unit vertikal Direktorat Jenderal Pajak menunjukkan fluktuasi jumlah wajib pajak terdaftar dari tahun ke tahun, sehingga diperlukan pendekatan analitis yang sistematis untuk mendukung perencanaan administrasi perpajakan yang bersifat prospektif. Penelitian ini bertujuan untuk meramalkan jumlah wajib pajak badan terdaftar di KPP Pratama Meulaboh menggunakan pendekatan deret waktu dengan model Autoregressive Integrated Moving Average (ARIMA). Data yang digunakan berupa data tahunan periode 1982–2024. Tahapan analisis meliputi transformasi data untuk menstabilkan varians, pengujian stasioneritas menggunakan uji Augmented Dickey–Fuller (ADF), identifikasi model melalui fungsi autokorelasi (ACF) dan autokorelasi parsial (PACF), serta pemilihan model berdasarkan kriteria Akaike Information Criterion (AIC). Hasil analisis menunjukkan bahwa model ARIMA(1,0,1) merupakan model terbaik dengan nilai AIC terendah sebesar –66,0104 dan parameter yang signifikan secara statistik. Uji diagnostik mengonfirmasi bahwa residual model memenuhi asumsi white noise. Model ARIMA(1,0,1) selanjutnya digunakan untuk menghasilkan peramalan jumlah wajib pajak badan untuk sepuluh tahun ke depan, yang menunjukkan tren peningkatan yang stabil sepanjang periode peramalan. Hasil penelitian ini memberikan dasar kuantitatif bagi otoritas pajak dalam perencanaan kapasitas administrasi, penguatan pengawasan kepatuhan, serta strategi perluasan basis pajak di wilayah kerja KPP Pratama Meulaboh.

Kata kunci: Deret Waktu; Peramalan; Model ARIMA; Wajib Pajak; Meulaboh

ABSTRACT

The number of registered taxpayers (WP) in a given region serves as a critical indicator of both the effectiveness of tax administration and the level of public participation in fulfilling tax obligations. KPP Pratama Meulaboh, as one of the vertical units of the Directorate General of Taxes, has experienced year-to-year fluctuations in taxpayer registration, highlighting the need for a systematic analytical approach to support forward-looking administrative planning. This study aims to forecast the number of registered corporate taxpayers at KPP Pratama Meulaboh using a time series approach based on the Autoregressive Integrated Moving Average (ARIMA) model. The dataset consists of annual observations covering the period 1982–2024. The analysis followed a structured procedure, including variance-stabilizing transformation, stationarity testing using the Augmented Dickey–Fuller (ADF) test, model identification through autocorrelation and partial autocorrelation analysis, and model selection based on the Akaike Information Criterion (AIC). Several ARIMA specifications were evaluated, and ARIMA(1,0,1) was selected as the

optimal model, yielding the lowest AIC value (–66.0104) and statistically significant parameters. Diagnostic checks confirmed that the model residuals satisfied the white noise assumption. The selected ARIMA(1,0,1) model was subsequently used to generate ten-year-ahead forecasts, which indicate a steady upward trend in the number of registered corporate taxpayers over the forecast horizon. These results provide practical insights for tax authorities in planning administrative capacity, strengthening compliance monitoring, and supporting strategic tax base expansion within the jurisdiction of KPP Pratama Meulaboh.
Keywords: *Time series; Forecasting; ARIMA model; Taxpayer registration; Meulaboh*

INTRODUCTION

Taxation serves as the backbone of state revenue, financing a wide range of development programs such as infrastructure, education, healthcare, and social subsidies. In developing countries like Indonesia, the role of taxation becomes increasingly vital as fiscal needs grow and reliance on foreign debt is no longer sustainable [1]. Consequently, optimizing tax revenue has become a central priority for the government, particularly through tax base expansion and the enhancement of taxpayer compliance.

One of the primary indicators of tax administration effectiveness is the number of officially registered taxpayers within the national tax system. An increasing number of taxpayers reflects successful tax extensification efforts, including initiatives such as public outreach, region-based monitoring, and cross-agency data integration [2]. Nevertheless, historical evidence suggests that taxpayer registration does not always follow a linear upward trend; fluctuations are often shaped by economic conditions, regulatory reforms, and shifts in public tax awareness [3].

KPP Pratama Meulaboh, a local office under the Directorate General of Taxes (DGT), plays a strategic role in collecting state revenue across West Aceh and its surrounding regions. As an office managing individual, corporate, and withholding taxpayers, KPP Pratama Meulaboh requires a comprehensive understanding of annual growth in registered taxpayers. Such information is not only essential for performance evaluation but also serves as a foundation for resource planning, budget allocation, and the design of more effective monitoring strategies [4].

To generate predictive insights into future taxpayer registration, quantitative approaches capable of capturing historical patterns and forecasting future trends are indispensable. Among these, time series analysis is widely recognized as a robust statistical method for modeling sequential data and producing reliable forecasts [5]. Within this framework, the Autoregressive Integrated Moving Average (ARIMA) model has gained prominence for its ability to handle non-stationary data through differencing and the integration of autoregressive and moving average components [6].

ARIMA models have been extensively applied across multiple disciplines, including economics, finance, and public management, due to their flexibility and strong predictive performance over short- to medium-term horizons [7]. Within the Indonesian empirical literature, ARIMA-based approaches have been employed to model various economic and financial time series, such as non-oil and gas export performance and stock price dynamics, demonstrating the model's robustness in capturing linear temporal patterns across diverse data contexts [10], [11]. Despite their methodological contributions, these studies primarily address macroeconomic and capital market indicators and do not engage with administrative-level forecasting problems in the taxation domain.

Previous studies have also demonstrated the applicability of ARIMA in taxation contexts, such as Nugroho and Ispriyanti [8], who focused on forecasting regional tax revenues, and [9], who applied ARIMA to predict the number of active taxpayers within a regional tax office in Central Java. Despite their methodological relevance, these studies are limited by relatively short observation periods and narrow analytical scopes, and they do not examine long-term registration dynamics using extended historical datasets. Moreover, empirical evidence on ARIMA-based forecasting of taxpayer registration at the level of individual tax offices in regions outside Java remains scarce. This gap motivates the present study, which utilizes a long-run annual dataset (1982–2024) to forecast taxpayer registration at KPP Pratama Meulaboh, thereby providing region-specific insights to support strategic tax administration planning.

Against this backdrop, the present study aims to forecast the number of registered taxpayers at KPP Pratama Meulaboh using the ARIMA model, based on annual data spanning 1982 to 2024. This study improves upon prior ARIMA-based tax forecasting research in several important aspects. First, it employs a substantially longer historical dataset, enabling the analysis of long-term registration dynamics rather than short-term fluctuations. Second, the analysis is conducted at the level of an individual tax office (KPP Pratama), providing a more granular and administratively relevant perspective than studies focusing on aggregate regional or provincial data. Third, by examining a non-Java region, this study contributes empirical evidence from an underrepresented geographic context in the Indonesian tax literature. The resulting forecasts are expected to provide reliable projections for the next decade, thereby offering a strategic foundation for improving monitoring activities, enhancing data collection, and strengthening taxpayer compliance within the jurisdiction of KPP Pratama Meulaboh.

METHOD

Data and Variables

This study employs a quantitative research design with a univariate time series forecasting approach. The objective is to model historical patterns in the number of registered taxpayers and to generate future projections using the Autoregressive Integrated Moving Average (ARIMA) model. The analysis is explanatory–predictive in nature, aiming to identify the underlying stochastic structure of taxpayer registration dynamics and to produce reliable forecasts that are relevant for administrative planning and policy formulation at the tax office level.

The data used in this study consist of annual observations of the total number of registered corporate taxpayers at KPP Pratama Meulaboh over the period 1982–2024 (see Table 1). The variable of interest is a single univariate time series representing taxpayer registration counts. The data were obtained from official administrative records of KPP Pratama Meulaboh and are assumed to be consistently recorded across years. The annual frequency enables the analysis of long-term registration dynamics and structural patterns in taxpayer growth.

Table 1 Registered Corporate Taxpayers at KPP Pratama Meulaboh, 1982–2024

Year	Registered Taxpayers	Year	Registered Taxpayers
1982	75	2004	505
1983	138	2005	364

1984	113	2006	1.019
1985	60	2007	3.502
1986	325	2008	2.951
1987	76	2009	14.311
1988	69	2010	7.065
1989	58	2011	6.225
1990	173	2012	3.628
1991	102	2013	4.328
1992	116	2014	4.617
1993	109	2015	6.398
1994	136	2016	5.940
1995	162	2017	5.112
1996	160	2018	8.137
1997	161	2019	8.333
1998	179	2020	27.165
1999	167	2021	6.474
2000	487	2022	8.092
2001	574	2023	11.723
2002	2.188	2024	12.262
2003	659		

Research Procedures

The research procedures were implemented through a structured and sequential workflow, as illustrated in Figure 1, which summarizes the complete ARIMA modeling process adopted in this study. The modeling procedure followed a systematic sequence comprising descriptive exploration, Box–Cox transformation for variance stabilization [12], Augmented Dickey–Fuller (ADF) stationarity testing [13], model identification based on the autocorrelation function (ACF) and partial autocorrelation function (PACF), parameter estimation, diagnostic checking, and forecasting [14].

As depicted in Figure 1, the analysis began with exploratory data analysis to examine long-term trends, variability, and potential structural patterns in the annual taxpayer registration series. To address possible heteroscedasticity, a Box–Cox transformation was applied when required to stabilize the variance prior to further modeling. Mean stationarity was then evaluated using the ADF test; non-stationary series were differenced iteratively until stationarity was achieved, consistent with standard time series modeling practices.

Following stationarity confirmation, candidate ARIMA models were identified based on the empirical behavior of the ACF and PACF plots. Parameter estimation was conducted using the maximum likelihood estimation method. Model adequacy was subsequently evaluated through residual diagnostic checking to ensure that the residuals satisfied white-noise assumptions, including independence and constant variance, as indicated in the decision nodes of Figure 1.

The general ARIMA (p, d, q) model can be expressed as:

$$\Phi\,(B)\,(1 - B)^{d}\,Y_t \;=\; \theta\,(B)\epsilon_t \tag{1}$$

where Y_t is the observation at time t, B is the backshift operator, $(1 - B)^d$ represents differencing, $\Phi\,(B)$ and $\theta\,(B)$ denote the autoregressive and moving average polynomials, respectively, and ϵ_t a white noise error term [13]. Model selection was guided by the Akaike Information Criterion (AIC), defined as:

$$AIC = -2ln(L) + 2k \tag{2}$$

where L is the likelihood function and k is the number of estimated parameters [14]. The model with the lowest AIC value and statistically significant coefficients was selected as the optimal forecasting model, as highlighted in the selection stage of Figure 1.

The finalized ARIMA model was then employed to generate ten-year-ahead forecasts of registered corporate taxpayers at KPP Pratama Meulaboh. This structured procedure, visually summarized in Figure 1, ensures methodological transparency and replicability of the forecasting framework.

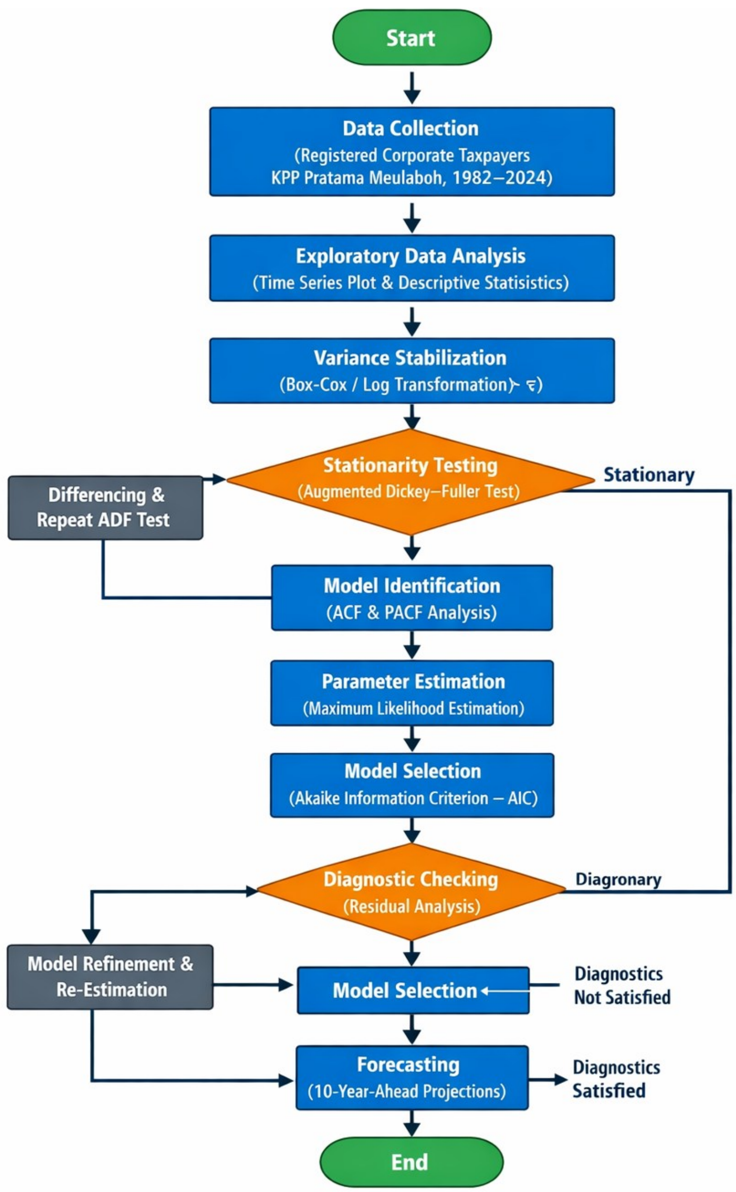

Figure 1. Flowchart Research Procedures

RESULT AND DISCUSSION

Stationarity Analysis

The first step in time series analysis is to assess the stationarity of the data. The dataset comprises the number of registered corporate taxpayers at KPP Pratama Meulaboh from 1982 to the most recent observation. The initial visualization in Figure 2 reveals an upward trend over time, indicating that the data are not fully stationary with respect to either the mean or the variance.

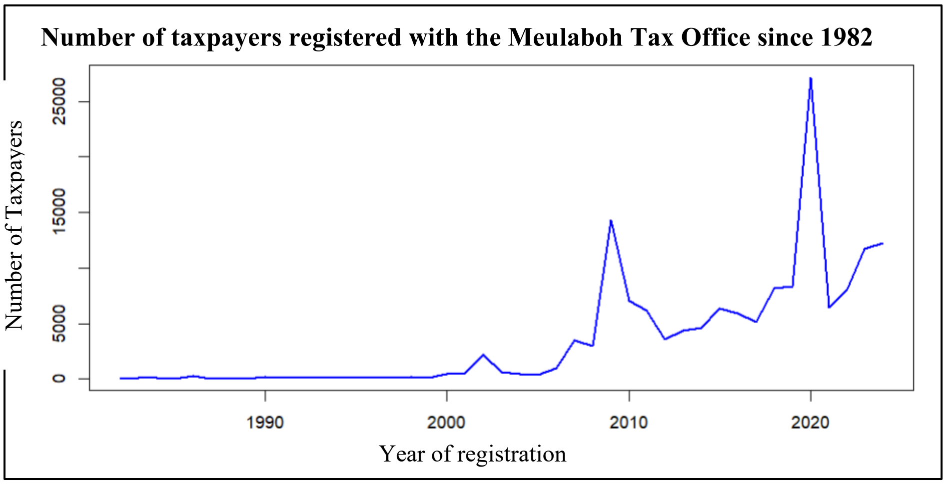

Figure 2. Time Series Plot of Registered Corporate Taxpayers

To evaluate variance stationarity, the Box–Cox transformation was applied to the data. The initial estimated lambda was 0.04985622, which deviates substantially from 1, suggesting unstable variance and the need for further transformation. A square-root transformation with an exponent of 0.17 was subsequently applied to stabilize the variance. After transformation, the recalculated lambda was 0.8869835, which is closer to 1, indicating improved homogeneity of variance. The visualization in Figure 3 confirms this result, showing more stable variance relative to the raw data, although the mean trend remains visible.

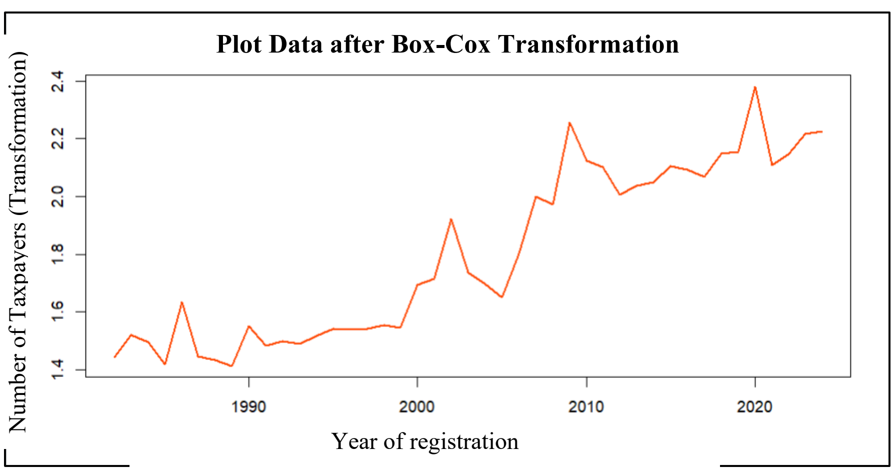

Figure 3. Time Series Plot After Box–Cox Transformation

To achieve mean stationarity, the transformed series was differenced once. The Augmented Dickey–Fuller (ADF) test applied to the differenced data yielded a Dickey–Fuller statistic of –4.6935 with a lag order of 3 and a p-value of 0.01. Since the p-value < 0.05, the null hypothesis of a unit root is rejected, indicating that the differenced series is stationary with respect to the mean. The resulting time series oscillates randomly around zero, as illustrated in Figure 4, further confirming stationarity.

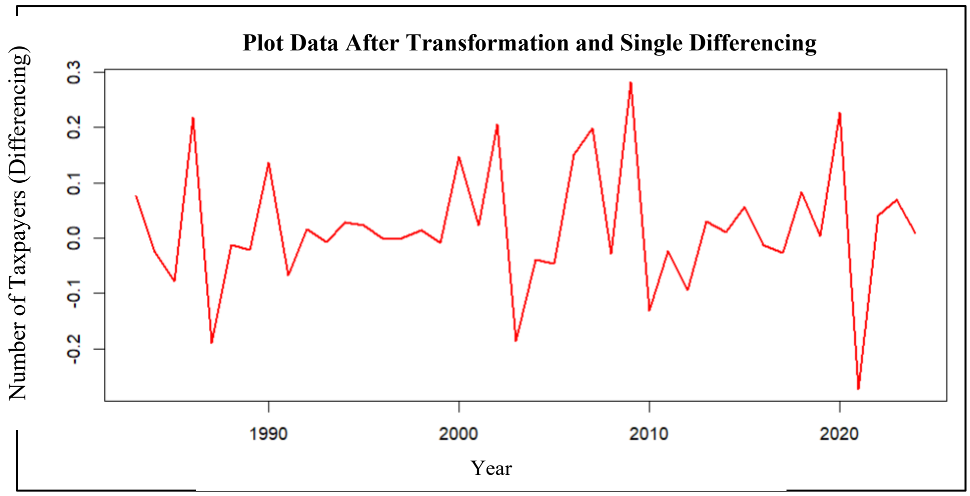

Figure 4. Time Series Plot After Box–Cox Transformation and First Differencing

To facilitate interpretation, the results of each transformation and differencing stage were summarized. The ADF test was conducted at three stages: the raw data, the Box–Cox transformed data, and the Box–Cox transformed data with first differencing. The summary in Table 2 indicates that both the raw data and Box–Cox transformed series exhibit p-values greater than 0.05, implying non-stationarity with respect to the mean. In contrast, after first differencing, the p-value fell below 0.05, allowing rejection of the null hypothesis and confirming stationarity.

Table 2. Stationarity Test Results for Registered Corporate Taxpayers

Testing Phase	Lambda Box-Cox	ADF Statistics	p-value	Conclusion
Raw data	0.0499	-	-	Variance unstable
After Box–Cox transformation	0.8870	- 2.5858	0.3423	Non-stationary with respect to the mean
After Box–Cox + 1st differencing	-	- 4.6935	0.0100*	Stationary with respect to mean and variance

These results indicate that a combination of variance transformation and first-order differencing was required to render the series of registered corporate taxpayers stationary before proceeding with ARIMA modeling.

Identification of ACF and PACF Patterns

Once the data achieved stationarity through transformation and first differencing, autocorrelation patterns were examined to identify potential ARIMA models. This analysis relied on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the transformed and differenced series. As illustrated in Figure 5, the ACF displayed a gradual decline (tailing off), while the PACF exhibited a sharp cut-off at specific lags. Such patterns typically suggest that the appropriate model is either an autoregressive (AR) process or a mixed ARMA structure.

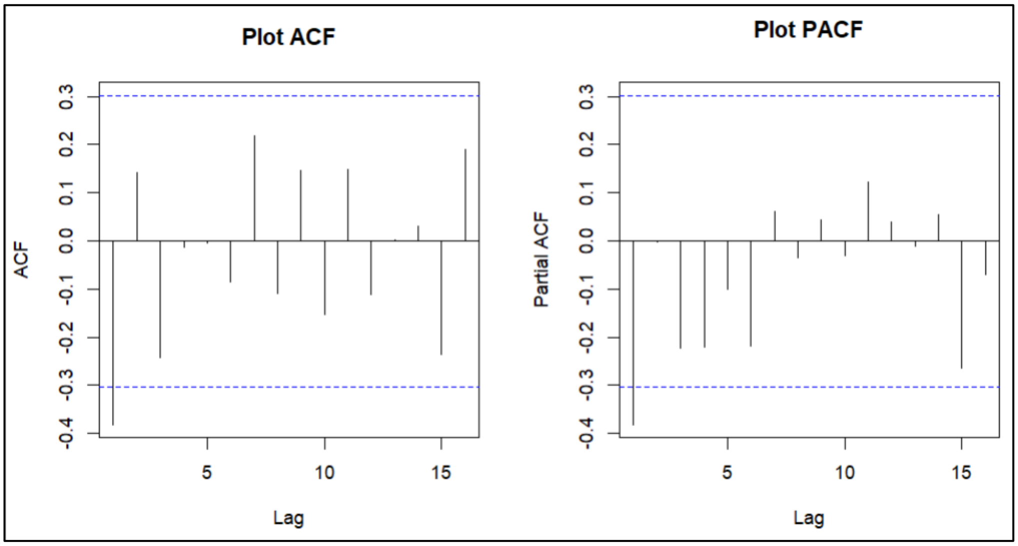

Figure 5. ACF and PACF Plots of Transformed and Differenced Data

Based on these ACF and PACF patterns, several ARIMA candidate models were evaluated using the stationary data to determine the best-fitting specification.

ARIMA Model Identification

Given the gradual decay observed in the ACF and the cut-off in the PACF (Figure 5), candidate models considered included AR(p) and ARMA(p,q). To ensure the selection of the most appropriate model, several ARIMA specifications were tested on the transformed and differenced series, namely ARIMA(1,0,1), ARIMA(3,0,1), ARIMA(3,0,0), ARIMA(0,0,1), ARIMA(1,0,0), ARIMA(1,0,2), and ARIMA(1,0,3). Each model was evaluated using the Akaike Information Criterion (AIC), with preference given to models balancing parsimony and fit (Table 3).

Table 3. Comparison of ARIMA Candidate Models

ARIMA Model	AIC Value
ARIMA(1,0,1)	-66,0104
ARIMA(3,0,1)	-64,02554
ARIMA(3,0,0)	-63,18401
ARIMA(0,0,1)	-65,95328

ARIMA(1,0,0)	-65,21406
ARIMA(1,0,2)	-64,88761
ARIMA(1,0,3)	-64,65981

The results indicated that ARIMA(1,0,1) yielded the lowest AIC, making it the optimal model for further evaluation. This model provided the best trade-off between accuracy and complexity. Subsequently, diagnostic checks were performed to validate model adequacy. Residual randomness was assessed using the Ljung-Box test, which confirmed the absence of significant autocorrelation ($p > 0.05$). Hence, the residuals of ARIMA(1,0,1) can be regarded as white noise, meeting the key assumptions of ARIMA modeling.

Forecasting the Number of Registered Taxpayers

Following the selection of ARIMA(1,0,1) as the best model, a ten-year forecast of the number of registered corporate taxpayers was generated. Forecasting was conducted on the differenced and transformed data, after which results were back-transformed to the original scale via inverse differencing and inverse transformation (Table 4).

Table 4. Ten-Year Forecast of Registered Corporate Taxpayers

Year	Forecast
2025	14.893
2026	17.354
2027	19.791
2028	22.313
2029	24.999
2030	27.906
2031	31.077
2032	34.551
2033	38.363
2034	42.550

The visualization of the forecast results is presented in Figure 5 below, which shows historical data and projections of the number of corporate taxpayers for the next 10 years.

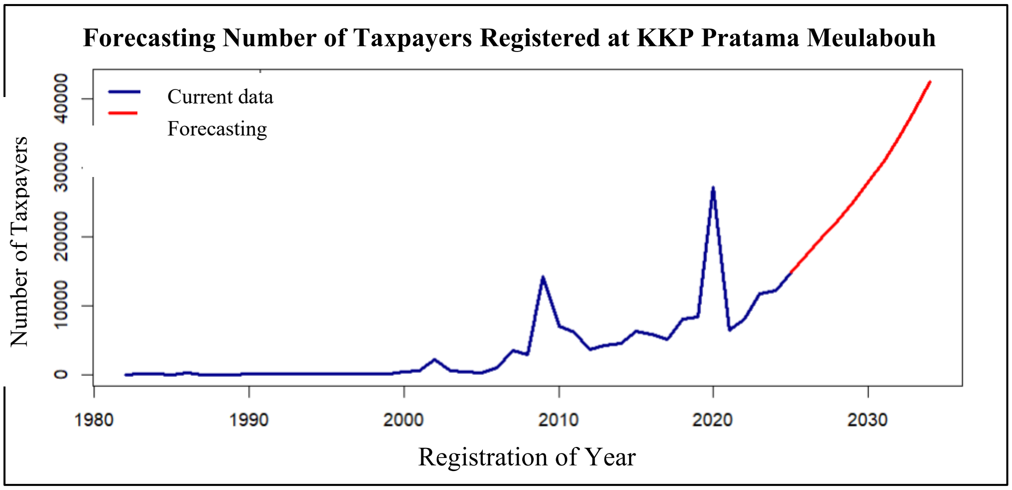

Figure 5. Historical Data and Forecast of Registered Corporate Taxpayers

The forecast results indicate a steady increase in the number of registered corporate taxpayers over the next decade, with projections surpassing 40,000 taxpayers by 2034. This upward trajectory is consistent with the historical pattern of gradual year-on-year growth observed in the registration data, suggesting a stable expansion of the taxpayer base within the jurisdiction of KPP Pratama Meulaboh.

From an administrative perspective, these projections carry important practical implications for tax management and planning. The anticipated increase in registered taxpayers implies a growing administrative workload related to taxpayer services, compliance monitoring, and data management. Consequently, the forecast results can be used as an evidence-based input for resource allocation, including staffing needs, audit capacity, and information system development at KPP Pratama Meulaboh.

Moreover, the projected growth provides a quantitative basis for strategic planning of taxpayer extensification and supervision programs. By anticipating the scale of future registration, tax authorities can better design outreach activities, prioritize compliance risk management, and align annual performance targets with realistic growth expectations. In this context, the ARIMA-based forecasts serve not only as statistical outputs but also as a planning instrument to support proactive and data-driven tax administration.

Overall, the forecasting results enhance the capacity of KPP Pratama Meulaboh to shift from reactive to forward-looking management, enabling more effective monitoring, improved service delivery, and strengthened taxpayer compliance over the medium term.

CONCLUSION

This study set out to forecast the number of registered corporate taxpayers at KPP Pratama Meulaboh using the ARIMA time series approach, and the results have met the expectations outlined in the Introduction. The analysis confirmed that the raw data were initially non-stationary, but after applying a Box-Cox transformation followed by first-order differencing, the series achieved stationarity and became suitable for modeling. Subsequent examination of the ACF and PACF patterns suggested the presence of both autoregressive and moving average components, and after evaluating several candidate models, ARIMA(1,0,1) was identified as the optimal specification. This model produced the lowest AIC value, exhibited statistically significant parameters, and generated residuals that satisfied the white-noise assumption, despite minor deviations from normality.

The forecasting exercise projected a steady increase in the number of registered corporate taxpayers over the next decade, rising from approximately 14.893 in 2025 to 42.550 in 2034. This sustained growth trend underscores the necessity for tax authorities, particularly KPP Pratama Meulaboh, to strengthen human resource capacity, adopt technology-driven innovations, and reinforce supporting infrastructure in order to accommodate the growing administrative burden. Recommendations emerging from this study emphasize the importance of periodic staff recruitment, continuous training, and the adoption of digital solutions such as e-filing, chatbots, and online queuing systems to enhance efficiency and service quality. At the same time, recognizing that ARIMA models rely exclusively on historical patterns, future research should integrate exogenous variables through approaches such as ARIMAX or VAR to better capture the effects of policy reforms, regional economic dynamics, and structural changes in the business sector.

Beyond its practical implications, this study contributes academically by demonstrating the applicability of ARIMA modeling in the field of fiscal administration, offering a quantitative framework that not only enhances the accuracy of taxpayer registration forecasts but also provides strategic insights for evidence-based decision-making in regional tax management.

REFERENCE

[1] L. L. Halawa, M. Sukma, E. Limbong, and W. Pangestoeti, “Efisiensi dan efektivitas pengelolaan penerimaan pajak untuk pembangunan ekonomi di Indonesia,” *Kajian Administrasi Publik dan Ilmu Komunikasi*, vol. 2, no. 2, pp. 258–269, 2025. https://doi.org/10.62383/kajian.v2i2.396

[2] Direktorat Jenderal Pajak, *Laporan Tahunan Direktorat Jenderal Pajak 2022*. Kementerian Keuangan Republik Indonesia, 2022.

[3] Organisation for Economic Co-operation and Development (OECD), *Tax Administration 2017: Comparative Information on OECD and Other Advanced and Emerging Economies*. OECD Publishing, 2017.

[4] Kementerian Keuangan Republik Indonesia, *Laporan Kinerja Instansi Pemerintah (LKjIP) Direktorat Jenderal Pajak Tahun 2023*. Kementerian Keuangan RI, 2023.

[5] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. OTexts, 2018.

[6] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley, 2015.

[7] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. Chapman & Hall/CRC, 2004.

[8] A. Nugroho and D. Ispriyanti, “Peramalan penerimaan pajak daerah menggunakan metode ARIMA,” *Jurnal Gaussian*, vol. 8, no. 2, pp. 175–184, 2019.

[9] F. N. Aini and T. Haryanto, “Proyeksi jumlah wajib pajak aktif menggunakan metode ARIMA,” *Jurnal Ekonomi dan Kebijakan Publik*, vol. 11, no. 1, pp. 45–56, 2020.

[10] M. I. Rizki, T. Ammar, F. Fitriyani, and S. Fasya, "Peramalan Indeks Harga Saham PT Verena Multi Finance Tbk Dengan Metode Pemodelan ARIMA Dan ARCH-GARCH", J Statistika, vol. 14, no. 1, pp. 11–23, Jul. 2021, doi: 10.36456/jstat.vol14.no1.a3774.

[11] F. N. Hayati, D. Nurlaily, and E. Pusporani, "Peramalan Data Ekspor Non Migas Provinsi Kalimantan Timur Menggunakan Univariate Time Series", J Statistika, vol. 14, no. 2, pp. 59–66, Jan. 2022, doi: 10.36456/jstat.vol14.no2.a3858.

[12] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–252, 1964.

[13] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427–431, 1979.

[14] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. Wiley, 2008.

[15] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. Springer, 2016.

[16] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.