International Journal of Mathematical Analysis and Modelling

(Formerly Journal of the Nigerian Society for Mathematical Biology)

Volume 8 Issue 2, 2025

ISSN (Print): 2682 - 5694
ISSN (Online): 2682 – 5708
www.tnsmb.org

Application of Markov models in income tax revenue forecasting in Nigeria

E. Adebiyi[*†] and **I. Adinya**[‡]

Abstract

Accurate tax revenue forecasting is vital for effective fiscal policy and economic planning, especially in contexts marked by data scarcity and economic volatility. Traditional statistical and machine learning models often underperform when confronted with non-linear patterns and limited historical data. This research proposes a novel hybrid Grey-Markov model that synergizes the trend-extraction strength of Grey System Theory (GM(1,1)) with the stochastic correction capability of Markov Chain Analysis. The Grey model forecasts long-term revenue trends using a minimal data requirement, while the Markov component dynamically adjusts predictions by modeling the probabilistic behavior of forecast errors. Through empirical validation using historical tax revenue data, the hybrid model will be benchmarked against conventional methods such as ARIMA, exponential smoothing, and neural networks. The anticipated outcome is a robust forecasting tool with improved accuracy and reliability, offering policymakers a more resilient framework for fiscal decision-making under uncertainty. This approach holds particular relevance for developing economies where data irregularities and economic shocks are prevalent.

Keywords and phrases: Grey model; Grey-Markov; income tax; forecasting policy;

1 Introduction

The income tax revenue of a nation is a critical indicator of its economic health and a fundamental component of its development. Taxation is pivotal for growth and development as it provides the essential funding for infrastructure and social amenities, contributing significantly to the well-being of the populace. Various tiers of government rely on tax revenue to fulfill their social obligations, with stable tax systems offering reliability that other revenue sources like levies and charges do not. There is no country in the world that completely operates without taxes, but there are countries with no personal income tax, such as the United Arab Emirates (UAE), Saudi Arabia, Qatar, etc. According to the IMF, developing countries must mobilize tax revenue to finance necessary services and infrastructure, the two being crucial for economic development. Consequently, accurate forecasting of tax revenue is indispensable for effective fiscal planning and economic stability.

Nigeria's current tax system faces numerous challenges, including the multiplicity of taxes, outdated legal provisions, and administrative inefficiencies, all of which hinder economic growth and discourage compliance. These issues disproportionately affect individuals and small businesses, especially the poor. Recognizing these challenges, ongoing tax reforms aim to simplify the tax structure, promote economic competitiveness, and enhance revenue generation in a fair and inclusive manner. However, the challenge of accurately forecasting tax revenue persists due to the complex and dynamic nature of economic activities. Nigeria, in particular, faces significant hurdles in tax collection and revenue prediction, which hamper effective fiscal planning and economic growth [3].

[*]Corresponding Author; E-mail: gift.adebiyi5@gmail.com
[†]Department of Mathematics, University of Ibadan, Ibadan, Nigeria
[‡]Department of Mathematics, University of Ibadan, Ibadan, Nigeria; E-mail: Iniadinya@gmail.com

Therefore, the implementation of advanced mathematical techniques, such as Markov models, is essential to improve the accuracy of income tax revenue forecasting in Nigeria. Markov models, known for their ability to handle random processes and state transitions, offer a robust framework for analyzing the stochastic nature of tax revenue generation, while the Grey Forecasting Model or abbreviated as GM (1,1) is a method used for forecasting a limited and short-term amount of data that produces an accurate forecasting model with no need to consider statistical distribution of data processed. Using historical data and statistical probabilities, these models can predict future revenue trends and help policymakers make informed decisions. This research aims to develop a comprehensive Markov model to assess the impact of various economic factors on income tax revenue in Nigeria, providing a valuable tool for strategic economic planning and policy formulation.
The innovative contribution of this paper lies in employing the GM-Markov combination model to predict income tax revenue from the formal employment sector in Nigeria, specifically between different consolidation of public service salaries (CONPSS) groups.

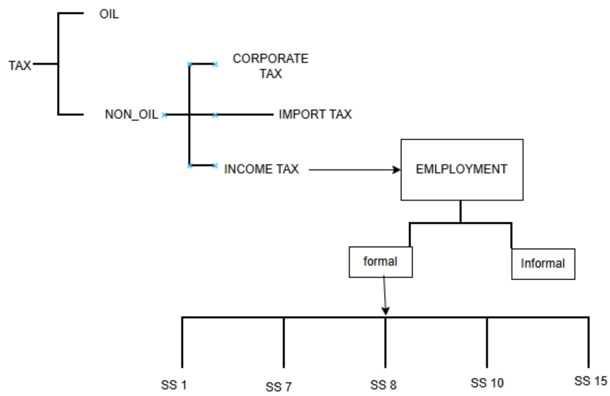


Figure 1: Caption for the tax model

By focusing on Consolidated Public Service Salary ranges in considering employment perspectives in Figure 1 below, the model can estimate how employees might move between salary scales (due to promotions, transfers, or career progressions) and calculate the resulting tax obligations dynamically. Salary structures are defined on the basis of the CONPSS scales[1], this model aims to offer a nuanced understanding of how different demographics contribute to tax revenue, facilitating more targeted and effective policy interventions. The GM-Markov combination model has been shown to be more accurate than single Grey prediction models in various contexts, such as predicting the income of rural residents in China [6]. By integrating the macro data of per capita disposable income over a period and using this combination model, researchers have achieved predictions closer to actual income figures and provided valuable policy suggestions for income growth. The formal employment sector in Nigeria, which accounts for approximately 10% of the workforce [4], is characterized by structured job roles, regular salaries, and benefits. The informal sector, while significant, presents numerous challenges, such as lack of regulation, variable income, limited data, and enforcement issues, making accurate forecasting difficult. Given these challenges, this research focuses on the formal sector due to its structured income data, reliable sources, and stronger policy impact. Despite its relatively small size compared to the informal sector, the formal sector plays a crucial role in the economy by providing stable employment and contributing significantly to the country's GDP [5].

Accurate forecasting of income tax revenue is vital for effective fiscal management in the Nigerian formal employment sector, particularly the civil services, which operate under the Consolidated Public Salary Structure (CONPSS). In a country with a diverse economic landscape, traditional forecasting methods, such as Artificial Neural Networks (ANN), Multiple Linear Regression (MLR), Seasonal

[1]Salary Scales and Ranges: CONPSS 1 Salary Scale: ₦930,000.00 – ₦1,002,621.00; CONPSS 7 Salary Scale: ₦1,277,667.00 – ₦1,653,093.00; CONPSS 8 Salary Scale: ₦1,479,276.00 – ₦1,914,514.00; CONPSS 10 Salary Scale: ₦1,806,041.00 – ₦2,356,897.00; CONPSS 15 Salary Scale: ₦3,014,528.00 – ₦3,847,474.00. [12].

Autoregressive Integrated Moving Average (SARIMA), and Holt-Winters, often fall short in capturing the complexities and nonlinearities present in tax data tied to tiered salary structures. This leads to significant discrepancies in revenue predictions.
To enhance forecasting accuracy, the adoption of advanced mathematical techniques such as the Grey Markov Model (GMM) is essential. The GMM combines the strengths of Grey System Theory and Markov Chains, making it particularly suitable for analyzing systems with small samples and uncertain information, such as tax revenue derived from the CONPSS salary scale.
The Grey System Theory component addresses the inherent uncertainties and trends in tax data, such as variations in employment levels, promotions, and economic factors affecting the CONPSS structure. Meanwhile, the Markov Chain component captures probabilistic state transitions across CONPSS levels and steps, such as movements due to promotions or step advancements. This dual approach ensures the model accommodates the nonlinear dynamics of the salary ranges, steps, and associated tax rates (e.g., 15% or 18%).
By explicitly integrating the CONPSS salary levels and steps into the model, along with the respective tax rates, the Grey Markov Model offers a robust framework for accurately predicting income tax revenue. It provides critical insights into the probabilistic behavior of tax revenue over time, accounting for transitions within salary levels, potential salary adjustments, and changes in employment trends.
This approach not only improves accuracy but also supports policy formulation, ensuring fiscal planning and public service funding are grounded in reliable, data-driven forecasts.

This study intends to bridge the gap in existing forecasting methods and contribute to the ongoing efforts of modernizing Nigeria's fiscal management practices [4] By analyzing the key determinants of tax revenue and applying advanced predictive techniques, this research offers practical insights and recommendations for enhancing the efficiency of Nigeria's tax system.
My motivation for this study stems from a desire to contribute to this Tax reform agenda by developing a more accurate and adaptable approach to tax revenue forecasting. By leveraging the predictive power of the Grey-Markov model, this research seeks to support evidence-based policymaking and improve the efficiency of fiscal planning in Nigeria. Accurate forecasts are essential for designing equitable tax policies that promote growth, ease the burden on the poor, and ultimately increase the tax-to-GDP ratio without compromising economic competitiveness.

Mao Zhan-li and Sun Jin-hua [8] introduced a Grey-Markov forecasting model to improve the prediction of fire accidents which is an area fraught with uncertainty and influenced by diverse fluctuating variables. Their approach cleverly integrates the GM(1,1) model, which excels with limited datasets, with the Markov chain, known for modeling randomness and state transitions. By first smoothing the data using the grey model and then capturing fluctuations through Markov analysis, they achieved a high level of forecasting accuracy. Testing their model on fire accident data from a Chinese city (2005–2009) demonstrated its superior performance over traditional models. Similarly, Muhammad Shodiq and colleagues [7] explored the role of forecasting in fisheries production, a critical sector for food security and economic sustainability. They highlighted the complexity of forecasting fish yields, which are influenced by environmental variables, fishing intensity, and human factors. Given the limited and often inconsistent data in this domain, they examined the effectiveness of various forecasting methods. Notably, their study found the GM(1,1) model to outperform traditional models like Single Moving Average (SMA) and Single Exponential Smoothing (SES), achieving an Average Relative Percentage Error (ARPE) of 9.60%, thereby proving its adaptability and accuracy. Despite the proven success of grey forecasting models and their hybrids in sectors with limited, noisy, or nonlinear data—such as fire safety and fisheries, these methods have not yet been applied to the domain of personal income tax forecasting. This represents a notable research gap, as personal income tax data often shares similar challenges: small sample sizes, volatility due to economic and behavioral factors, and limited historical consistency. Traditional statistical models may fail to accommodate these complexities, leading to suboptimal predictions and policy planning.
Huidi Zhang and Yimao Chen [17] explore the complexities of predicting tax revenues influenced by

dynamic economic and political factors. They critique traditional forecasting methods for their limitations and propose the Grey-Markov chain model, which combines grey forecasting and Markov chain forecasting. Their simulation using tax revenue data from a Chinese province shows that this model improves forecasting accuracy by 2.3% to 13.1% compared to traditional methods. The study highlights the model's practical application in analyzing tax trends and its effectiveness in providing reliable predictions, making it a valuable tool for policymakers and researchers.
Inyeneobong Edem et al [[10]] Introduce an innovative forecasting tool designed to improve the prediction of fire accidents in manufacturing and processing industries. The modified grey-Markov model (MMSGM) utilizes information turbulence indices (ITI) and restricted residuals to enhance accuracy, particularly when historical data is limited. Validation through both simulated scenarios and real-time applications demonstrates the model's effectiveness, significantly outperforming traditional grey and grey-Markov models in terms of forecasting precision. However, this study presents several gaps that warrant further exploration. While the MMSGM shows promise in the oil and gas sector, its applicability across diverse industries remains untested, limiting the generalizability of the findings. Additionally, the model's reliance on historical data being within a specific ITI range (not exceeding 0.3) raises questions about its robustness in varying conditions. The authors also acknowledge a slight decline in computational efficiency compared to other forecasting models, suggesting that future research could focus on optimizing performance without compromising accuracy. Furthermore, the absence of external validation through industry benchmarking indicates a need for broader application of the MMSGM to strengthen its credibility. Lastly, the study could benefit from further discussion on integrating emerging technologies, such as machine learning, to enhance forecasting capabilities.

D'Amico, Karagrigoriou, and Vigna [[18]] present a pioneering approach to predicting Italy's energy landscape. By integrating Grey system theory with Markov models, they tackle the complexities inherent in energy forecasting. Their analysis, which spans data from 2000 to 2022, reveals a remarkable enhancement in accuracy, achieving over a 50% reduction in the Root Mean Squared Error (RMSE).

However, while the study shines a light on the anticipated growth of renewable energy sources in Italy, it also uncovers a persistent reliance on fossil fuels, predicting that they will continue to dominate the energy mix by 2032. This critical finding underscores the urgent need for policymakers to implement stronger measures to accelerate the transition to renewable energy and reduce dependence on fossil fuels.

Yet, the paper leaves several gaps that invite further exploration. For instance, the projections are confined to the year 2032, leaving us in the dark about the long-term trajectory of Italy's energy mix beyond this point. Additionally, the study does not consider how evolving variable factors—such as technological advancements, economic fluctuations, and changes in policy—might influence future energy scenarios.

The analysis also overlooks regional variations within Italy, which could play a significant role in shaping the energy landscape. Lastly, the authors do not delve into the importance of stakeholder engagement in the development of energy policies, an element that is crucial for crafting effective strategies.

These identified gaps highlight opportunities for further research that could not only enrich the insights provided by this study but also contribute to a deeper understanding of Italy's energy transition landscape.

Samuel O. [[16]] Utilizing secondary data from the NNPC's 2019 annual statistics bulletin, the authors demonstrate that the MGM(1,1, b) model provides more accurate predictions for natural gas consumption from 2021 to 2025 compared to the traditional GM(1,1) model. The results indicate that the predictions fall within a defined interval, offering valuable insights for policymakers and stakeholders in Nigeria's energy sector.

However, several gaps remain in the research. Firstly, while the study focuses on a short-term forecast, it does not explore the potential for extending the model to long-term predictions, which is crucial for strategic energy planning. Additionally, the model does not account for external factors such as economic fluctuations or policy changes that could significantly influence natural gas consumption.

Nguyen et al [[13]] use data from 2014 to 2019, the study forecasts unemployment rates for 2020 to 2025 across six regions in Vietnam. The findings reveal that the GM(1,1) model consistently outperforms the others, achieving lower Mean Average Percentage Error (MAPE) values, indicating its effectiveness in capturing the dynamics of unemployment in the country. However, several gaps remain in the research. The study does not explore long-term forecasting beyond 2025 or account for external factors like economic policies that could influence unemployment rates. Additionally, while it compares Grey-based models with ARIMA, it lacks a broader analysis with other forecasting methods, such as machine learning techniques

Yi-Chung and Jiang [[19]] highlight the limitations of traditional forecasting methods, particularly in handling fluctuating time series data. They advocate for the Grey–Markov model as a more robust alternative to the GM(1,1) model, especially in volatile conditions. By incorporating two neural networks (NNs), the study addresses common challenges, such as the need for background value determination and the adjustment of predicted values.

Empirical validation using historical data from the Taiwan Tourism Bureau and the China National Tourism Administration demonstrates that the NN-MCRGM(1,1) model significantly outperforms traditional Grey–Markov models [[15]], showcasing its effectiveness in capturing tourism demand complexities.

Huda et al [[15]] introduce Grey System theory, developed by Deng Julong in 1982, highlighting its advantages in scenarios with limited or unstructured data. By integrating Markov Chain analysis, they enhance the Grey model into the Grey-Markov (1,1) model, which allows for state transitions over time. Using data from the Central Statistics Agency covering 1996 to 2021, they employ a systematic approach that includes the Accumulated Generating Operation (AGO) and the Mean Generating Operation (MGO) to forecast health complaint percentages. Their findings predict a 30.36% rate for 2022, with an accuracy of 2.43%. Huda and Imro'ah's research provides valuable insights into the Grey-Markov (1,1) model's application for forecasting health complaints in Indonesia. Addressing the identified gaps could enhance the model's effectiveness and its utility in public health forecasting.

Lawal et al [[14]] This study explores the use of the Grey system model (GM(1,1)) and the Grey-Markov model to forecast Nigeria's annual rice production using data from 2010 to 2015. The models demonstrated high accuracy, with a mean absolute percentage error (MAPE) of 1.02%. The research highlights the importance of reliable crop yield forecasts for stakeholders in agriculture, emphasizing the need for accurate predictions to inform government policies and food security plans. The findings suggest that these models can significantly aid in sustainable rice production in Nigeria. The research also overlooks regional variability in rice production across Nigeria, which is essential given the country's diverse agricultural practices. Engaging stakeholders in the research process and discussing how the findings can be utilized would further enhance the practical applicability of the forecasts. Lastly, while the study provides short-term predictions, it does not explore long-term forecasting, which is vital for strategic planning in food security.
Despite these strengths, the study identifies several gaps for future research. The model's performance in volatile datasets remains untested, and external factors influencing health complaints are not considered. Additionally, the interpretability of predictions is not addressed, which is crucial for stakeholders. The focus on short-term forecasting limits the model's applicability, and comparative analyses with other forecasting methods are lacking.

Despite these advancements, several gaps remain. The model's applicability to other geographical contexts is unexplored, limiting its generalizability. Additionally, the study does not consider external factors influencing tourism demand, such as economic conditions or global events. The interpretability of neural network outputs is also not addressed, which is crucial for stakeholders relying on these forecasts. Lastly, the potential limitations of the Grey–Markov model itself warrant further examination.

In conclusion, while the integration of neural networks into Grey–Markov models represents a significant advancement in tourism demand forecasting, addressing these gaps in future research could enhance the model's applicability and effectiveness. Therefore, adapting and applying Grey-Markov models to the forecasting of personal income tax revenues could offer a novel, data-efficient, and dynamic approach. This would not only bridge a methodological gap but also provide tax authorities and policymakers with more accurate tools for fiscal planning, revenue estimation, and economic management, especially in developing economies where tax data is scarce or inconsistent, like in Nigeria.

2 Methodology/Model formulation

2.1 Mathematical model

To build a mathematical framework for forecasting income tax revenue using the Grey-Markov model, we will combine two key components: **Markov Chains** and **Grey Models**. This approach leverages the strengths of both models to predict future income tax revenue with a focus on structured transitions and adaptability to sparse or uncertain data.

2.1.1 Markov chain model for income tax revenue

Markov models are effective for systems with probabilistic state transitions. In this case, we can treat tax revenue from various demographics or sectors as "states" within a Markov chain framework. The Markov model will help analyze the transition probabilities between different revenue states over time.

State space definition

Define a discrete set of states for tax revenue, each representing a specific revenue level or income group, i.e it represent the proportion of employees in each CONPSS salary level, rather than tax revenue directly. Each state corresponds to the distribution or probability of an employee being within a specific salary tier at time t.

$$S = \{s_1, s_2, \ldots, s_n\}$$

Each state s_i represents a revenue level or income group. For instance, if focusing on groups, s_1 could represent tax revenue from the s_1 in CONPSS 1, s_2 in CONPSS 7, and so forth.

$$S = \{s_1, s_2, s_3, s_4, s_5\}$$

s_1 : Tax revenue from the CONPSS 1
s_2 : Tax revenue from the CONPSS 7
s_3 : Tax revenue from the CONPSS 8
s_4 : Tax revenue from the CONPSS 10
s_5 : Tax revenue from the CONPSS 15

This state space explicitly accounts for the realistic movement of employees not only among salary tiers but also their exit from the employment system, which is critical to modeling workforce dynamics. Tax revenue from each salary level i at time t is computed as:

$$R_i(t) = \pi_i(t) \cdot N(t) \cdot r_i(t) \tag{1}$$

where:

- $\pi_i(t)$: proportion of employees in state s_i at time t
- $N(t)$: total number of formal sector employees at time t
- $r_i(t)$: effective tax rate for CONPSS level i at time t

Total tax revenue is then:

$$R(t) = \sum_{i=1}^{5} \pi_i(t) \cdot N(t) \cdot r_i(t) \tag{2}$$

Transition probabilities

Let P_{ij} represent the probability of transitioning from state s_i to s_j. These probabilities are structured in a **transition matrix** P:

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}$$

where:

$$P_{ij} = \Pr(\text{next state } = s_j \mid \text{current state } = s_i), \quad \sum_{j=1}^{n} P_{ij} = 1$$

The matrix P provides insights into the likelihood of shifting between revenue levels or age groups, capturing the dynamic nature of tax revenue changes over time.

Let P_{ij} represent the probability of transitioning from state s_i to s_j. These probabilities are structured in a **transition matrix** P:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} \\ P_{21} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} & P_{36} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} & P_{46} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} & P_{56} \\ RDS & RDS & RDS & RDS & RDS & RDS \end{bmatrix}$$

where P_{ij} represents the probability of transitioning from state s_i to state s_j, and the sum of each row's elements is 1:

$$\sum_{j=1}^{6} P_{ij} = 1 \quad \text{for each } i = 1, 2, \dots, 6.$$

2.1.2 Grey model for income tax revenue

Grey forecasting model is a set of mathematical models given by differential equations which can be used to observe a single and predict the development and changes of the system under study. The advantage of Grey forecasting lies in short-term forecasting but the disadvantage lies in its poor ability of long-term forecasts and volatile data series.

Whereas Markov model enhances income tax by allowing for more accurate predictions of future tax revenue based on historical data.

By analyzing past income data, a Markov model can indicate the probability of an individual moving between different income brackets in the future, which helps in forecasting overall tax revenue.

Also by observing patterns of income changes the model can identify taxpayers with a higher likelihood of significant income fluctuations, potentially signifying fraudulent activity.

Likewise, Markov models can be used to simulate the impact of potential tax policy changes on future tax revenue, enabling policymakers to assess the effectiveness of different strategies.

Lastly, by utilizing historical data, the model can reveal hidden trends in taxpayer behavior that might not be apparent through traditional analysis methods. However,Markov models assume that future state depends only on the current state, which may not fully capture the complexity of real-world economic factors affecting taxpayer income.

Hence, the combination of Grey and Markov model helps to reveal the general trends of the development and change of the data sequence and predict the state of the data values making the forecasting more scientific and practical.

In using the Grey Markov Model, We first establish a Grey theory system and find the free casting curve. Then use the smooth free casting curve as a benchmark to divide a number of dynamic state intervals. Calculate the Markov transition probability matrix at the points and predict its future state. Get the predicted value interval and take the midpoint of the interval, and finally get the predicted value with higher accuracy Grey Model (GM(1,1))

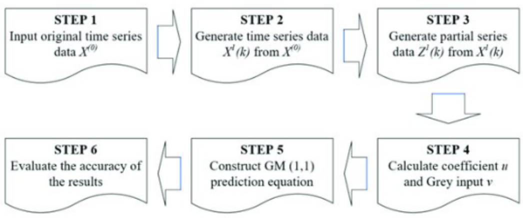


Figure 2: (Source: Research Gate): The Grey Model GM(1,1) process

2.2 Baseline model

Consider the new Nigeria Tax Bill Reform proposed by the President exempts individuals earning ₦50 million or less annually from paying personal income tax. However, individuals earning above ₦50 million will be subject to a 24% personal income tax rate. This reform raises questions about its impact on various income groups, particularly civil servants on Grade Level 1.

As at the year 2022, the total population of Grade Level 1 civil servants in Nigeria was reported to be 720,000. By June 2022, this population had reduced to 513,000, according to Insight on Government UK.

- **Exempt Group:** Civil servants earning ₦800 thousand or less annually fall within the tax-exempt category under the new tax reform.
- **Taxable Group:** Any individual earning above ₦50 million annually will be required to pay 24% of their income as personal income tax.

CONPSS Level	Education	Steps	Salary Range (₦)	Tax
CONPSS 1	O'Level	1-14	₦930,000 – ₦1,002,621	15%
CONPSS 7	OND/NCE	1-14	₦1,277,667 – ₦1,653,093	15%
CONPSS 8	BSc	1-14	₦1,479,276 – ₦1,914,514	15%
CONPSS 10	Professional Qualification	1-14	₦1,806,041 – ₦2,356,897	15%
CONPSS 15	Appointment Post	1-9	₦3,014,528 – ₦3,847,474	18%

Table 1: Alignment of Proposed Tax Rates with CONPSS Salary Scales [20]

By focusing on age ranges, this model aims to offer a nuanced understanding of how different demographics contribute to tax revenue, facilitating more targeted and effective policy interventions. With the adoption of Markov models, Nigeria can improve its revenue forecasting accuracy, thereby strengthening its economic stability and fostering sustainable growth.
for easy calculation

$$\text{Average salary} \approx \frac{\text{Max Salary} + \text{Min Salary}}{2}$$

CONPSS	Education	Steps	Average Salary (₦)	Tax
CONPSS 1	O'Level	1-14	₦970000	15%
CONPSS 7	OND/NCE	1-14	₦1,500,000	15%
CONPSS 8	BSc	1-14	₦1,700,000	15%
CONPSS 10	Professional Qualification	1-14	₦2,100,000	15%
CONPSS 15	Appointment Post	1-9	₦3,500,000	18%

Table 2: Constant Intervals and Tax Rates for CONPSS Levels

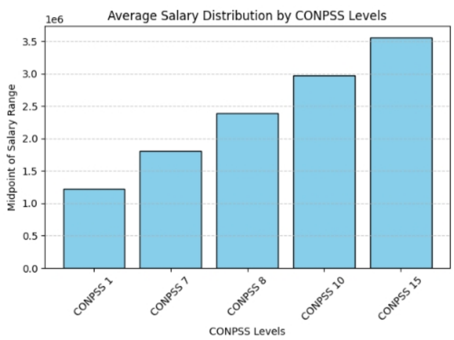


Figure 3: Average salary Distribution by CONPSS Levels

2.2.1 Reasons for using average salaries

The adjustment from the salary ranges in the first table to the simplified average salaries in the second table reflects the following considerations:

1. Simplification of Data: Calculating tax using a single average value for each level is easier and more consistent than dealing with ranges. This simplifies computations, especially for broader analysis or projections.

2. Fair Representation of Salary Ranges: The average salary provides a central value that represents the midpoint of the salary range for each CONPSS level, capturing the typical earning within the level

3. Practicality in Reporting: Many financial and tax-related models require a single figure for each category to avoid complexities in handling ranges. Using an average salary standardizes the representation.

4. Uniform Basis for Tax Calculation: Since the tax rates (15% or 18%) apply uniformly to the entire salary range, calculating based on the average salary provides a consistent approximation of the total tax impact for employees in each CONPSS level.

2.3 Modeling the Assumptions

In applying Grey-Markov model to the forecasting of the CONPSS income tax generation, two assumptions will be made: Model based on individual/employee with Certificate and Model based on individual/employee with 35 years of services without further education.

2.3.1 Transition Matrix for an Individual with Certificate:

Considering grade levels 1, 7, 8, 10, and 15, we made the following assumptions:

- Once in a grade level, you can only move to upper levels or go to Retirement/Death/Sack (R.D.S.) end.
- Grade level 1 can only move to grade level 7 or 8 but not the rest of the grade levels because it requires a degree certificate to jump to grade level 10 or 15.

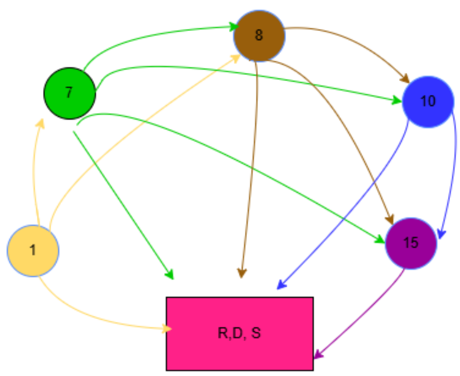


Figure 4: Transition Matrix for an Individual with Certificate

The loop from one state to another (grade level) is mainly by obtaining an education certificate and experience needed for that grade level, except for grade level 15, which is by appointment.

The loop from one state to another is represented by arrows. The arrows represent the transition from one state to another. A node represents each state.

i represents Grade level 1
j represents Grade level 7
k represents Grade level 8
l represents Grade level 10
m represents Grade level 15
n represents R, D, S state (Retirement/Death/Sack)

The matrix will be in the following form:

	i	j	k	l	m	n
i	ii	ij	ik	il	im	in
j	ji	jj	jk	jl	jm	jn
k	ki	kj	kk	kl	km	kn
l	li	lj	lk	ll	lm	ln
m	mi	mj	mk	ml	mm	mn

Assumption (n):

$$
\begin{array}{|cccccc|c|}
\hline
 & & & & & & \text{R, D, S} \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\hline
\end{array}
=
\left[\begin{array}{cccccc|c}
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{array}\right]
$$

The individual/employee with the certificate, form a trivial formulation which provides no forecasting value and will be been excluded from the model to maintain methodological rigor and scientific credibility.

2.3.2 Transition Matrix for an Individual with 35 Years of Service without Further Education:

Transition matrix model assumptions

- We assume Individuals will join the civil services without a degree certificate
- We assume that the individual progresses through the steps in each level, with each step taking on average three years [9].

Using the above figure, the transition matrix below shows the probabilities of moving from one level to the next annually:

Where:

- $P(\text{CONPSS 1 to CONPSS 7}) = \frac{14}{35} = 0.4$
- $P(\text{CONPSS 1 to CONPSS 7 to CONPSS 8}) = 0.4 \times 0.4 = 0.16$
- $P(\text{CONPSS 1 to CONPSS 7 to CONPSS 8 to COMPSS 10}) = 0.4 \times 0.4 \times 0.4 = 0.064$
- $P(\text{CONPSS 1 to CONPSS 7 to CONPSS 8 to COMPSS 10 to CONPSS 15 }) = 0.4 \times 0.4 \times 0.4 \times 0.257 = 0.02$

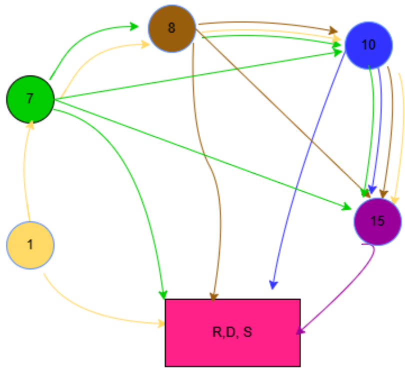


Figure 5: Transition Matrix for an Individual without Further Education

- P(CONPSS 1 to CONPSS 7 to CONPSS 8 to COMPSS 10) $= 0.4 \times 0.4 \times 0.4 = 0.16$
- P(CONPSS 7 to CONPSS 8) $= \frac{14}{35} = 0.4$
- P(CONPSS 7 to CONPSS 8 to CONPSS 10) $= 0.4 \times 0.4 \times 0.257 = 0.04$
- P(CONPSS 7 to CONPSS 8 to CONPSS 10 to CONPSS 15) $= 0.4 \times 0.4 \times 0.257 = 0.04$
- P(CONPSS 8 to CONPSS 10) $= \frac{14}{35} = 0.4$
- P(CONPSS 10 to CONPSS 15) $= \frac{9}{35} = 0.2571$
- P(CONPSS 7 to CONPSS 10) $= \frac{14}{35} = 0.4$
- P(CONPSS 7 to CONPSS 1 CONPSS 15) $= \frac{9}{35} = 0.2571$
- P(CONPSS 8 to CONPSS 10 to CONPSS 15) $= 0.4 \times 0.2571 = 0.1$
- P(CONPSS 8 to CONPSS 15) $= \frac{9}{35} = 0.2571$
- P(CONPSS 1 to RDS) $= 1$
- P(CONPSS 7 to RDS) $= 1$
- P(CONPSS 8 to RDS) $= 1$
- P(CONPSS 10 to RDS) $= 1$
- P(CONPSS 15 to RDS) $= 1$

The matrix will be in the following form:

$$\begin{array}{c|cccccc}
 & i & j & k & l & m & n \\
\hline
i & ii & ij & ik & il & im & in \\
j & ji & jj & jk & jl & jm & jn \\
k & ki & kj & kk & kl & km & kn \\
l & li & lj & lk & ll & lm & ln \\
m & mi & mj & mk & ml & mm & mn
\end{array}$$

The transition matrix is:

$$\left[\begin{array}{ccccc|c} 0.356 & 0.4 & 0.16 & 0.064 & 0.02 & 1 \\ 0 & 0.14 & 0.4 & 0.16 & 0.30 & 1 \\ 0 & 0 & 0.24 & 0.4 & 0.36 & 1 \\ 0 & 0 & 0 & 0.7 & 0.3 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

2.4 Grey model for trend forecasting

The GM-Markov model integrates the strengths of Grey and Markov models to account for both long-term trends and short-term fluctuations in tax revenue. The **Grey Model (GM(1,1))** is a first-order Grey model that predicts a single-variable time series.

Given a time series $x = \{x(1), x(2), \ldots, x(n)\}$, we denote the original sequence as $x^{(0)}$.

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, 2, \ldots, n \tag{3}$$

$$\frac{dx^{(1)}(k)}{dk} + ax^{(1)}(k) = b. \tag{4}$$

where a is a coefficient that represents the decay rate of $x^{(1)}(t)$. it represent the sensitivity of tax revenue to economic growth or contraction, reflecting how quickly the revenue adapts to broader economic changes, like shifts in employment, business activity, or inflation. and the Paramenter b is a constant term that can represent a driving or input term in the system i.e baseline revenue level or a driver representing external influences that consistently affect income tax revenue, such as government fiscal policies, foreign investments, or social welfare programs. The given first-order differential equation is:

Given the original time series: $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$

AGO: $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$

from the mean sequence; $z^{(1)}(k) = \theta x^{(1)}(k) + (1-\theta)x^{(1)}(k-1)$, typically $\theta = 0.5$

where:

- a: development coefficient (negative, captures trend decay/growth)
- b: driving coefficient (data driving force)

Using the equation is first-order linear.
$\frac{dx^{(1)}}{dt} + P(t)x^{(1)} = Q(t)$, where $P(t) = a$, $Q(t) = b$ (constants).

Integrating factor: $e^{\int a\,dt} = e^{at}$

Multiply through:

$$e^{at}\frac{dx^{(1)}}{dt} + ae^{at}x^{(1)} = be^{at} \tag{5}$$

Left side: $\frac{d}{dt}\left[x^{(1)}e^{at}\right]$

$$\frac{d}{dt}\left[x^{(1)}e^{at}\right] = be^{at} \tag{6}$$

Integrate both sides (from $t = 1$ to $t = k$):

$$
\begin{aligned}
x^{(1)}(k)e^{ak} - x^{(1)}(1)e^{a} &= b\int_1^k e^{at}\,dt \\
\int_1^k e^{at}\,dt &= \frac{1}{a}\left[e^{ak} - e^{a}\right] \\
x^{(1)}(k)e^{ak} &= x^{(1)}(1)e^{a} + \frac{b}{a}(e^{ak} - e^{a}) \\
x^{(1)}(k) &= x^{(1)}(1)e^{a(1-k)} + \frac{b}{a}(1 - e^{a(1-k)})
\end{aligned} \tag{7}
$$

For **discrete forecasting** at integer steps $k + 1$:

$$\hat{x}^{(1)}(k+1) = \left(x^{(1)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a} \tag{8}$$

Proof of form: Let $C = x^{(1)}(1) - \frac{b}{a}$, then:

$$\hat{x}^{(1)}(k+1) = Ce^{-ak} + \frac{b}{a}$$

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{9}$$

Substitute (8):

$$\hat{x}^{(0)}(k+1) = (1 - e^{a})\left(x^{(1)}(1) - \frac{b}{a}\right)e^{-ak} \tag{10}$$

This exact solution enables the Grey component of the Grey-Markov-CONPSS tax forecasting model, providing the trend component before Markov state transition corrections.

$$x^{(0)}(k) + az^{(1)}(k) = b.$$

which can be express as a linear regression formula

$$y = B\hat{a} + \epsilon \tag{11}$$

whose least squares estimation formula can be expressed:

$$\hat{a} = (B^TB)^{-1}B^Ty$$

where:

- y is the vector of observed values,
- B is the design matrix containing the independent variables (in our case, the $x^{(1)}$ values and a column for the intercept),
- $\hat{a}$ is the vector of parameters we want to estimate,
- ϵ is the error term.

The goal of least squares is to minimize the sum of squared residuals (the differences between observed values and predicted values) [see [11]]:

$$S = \sum (y_i - (B\hat{a})_i)^2$$

Solve for a and b using the least squares method: The solution associated with (11) is given by

$$\begin{bmatrix} a \\ b \end{bmatrix} = \left(B^T B\right)^{-1} B^T Y$$

where:

$$B = \begin{bmatrix} -0.5\left(x^{(1)}(1) + x^{(1)}(2)\right) & 1 \\ -0.5\left(x^{(1)}(2) + x^{(1)}(3)\right) & 1 \\ \vdots & \vdots \\ -0.5\left(x^{(1)}(n-1) + x^{(1)}(n)\right) & 1 \end{bmatrix}$$

and

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

The solution to the differential equation provides the forecasted values $\hat{x}^{(1)}(k)$:

$$\hat{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-a(k-1)} + \frac{b}{a}$$

To obtain the forecasted values in the original series $\hat{x}^{(0)}(k)$, use the Inverse Accumulated Growth Over (IAGO):

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1)$$

for $k = 2, 3, \ldots, n$.

Forecast the distribution across states over time using:

$$\pi(t) = \pi(0) \cdot P^t$$

The forecasted state distribution $\pi(t)$ at time t can be written as:

$$\pi(t) = \begin{bmatrix} \pi_1(0) & \pi_2(0) & \ldots & \pi_n(0) \end{bmatrix} \cdot \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1n} \\ P_{21} & P_{22} & \ldots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \ldots & P_{nn} \end{bmatrix}^t$$

The forecasted tax revenue for state s_i at time t is:

$$R_i(t) = \pi_i(t) \cdot N(t) \cdot r_i(t)$$

The total forecasted tax revenue at time t is:

$$R(t) = \sum_{i=1}^{n} R_i(t) = \sum_{i=1}^{n} \pi_i(t) \cdot N(t) \cdot r_i(t)$$

The total forecasted tax revenue $R(t)$ is the sum of the revenues from all states $s_1, s_2, \ldots, s_n$. Each state's contribution to the total is based on its proportion of the total revenue $\pi_i(t)$, the number of individuals $N(t)$, and the tax rate $r_i(t)$. The formula for the total forecasted tax revenue is:

$$R(t) = \sum_{i=1}^{n} R_i(t) = \sum_{i=1}^{n} \pi_i(t) \cdot N(t) \cdot r_i(t)$$

Where: - $R(t)$ is the total forecasted tax revenue at time t, - $\pi_i(t)$ is the proportion of the total revenue coming from state s_i at time t, - $N(t)$ is the total number of individuals contributing to the tax revenue at time t, - $r_i(t)$ is the tax rate or average revenue generated per individual in state s_i at time t.

2.5 Derivation of the Grey–Markov Forecast

Let the original non–negative series be

$$x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n).$$

Apply the 1–AGO (accumulated generating operation):

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, \ldots, n.$$

For the GM(1,1) model, the whitening differential equation is

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b,$$

whose solution with initial condition $x^{(1)}(1)$ is

$$\hat{x}^{(1)}(k+1) = \big(x^{(1)}(1) - \tfrac{b}{a}\big)e^{-ak} + \tfrac{b}{a}, \quad k = 0, 1, \ldots$$

The restored forecast for the original sequence is obtained by

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad k = 1, 2, \ldots$$

The values $\hat{x}^{(0)}(k)$ are referred to as the Grey forecasts.

To incorporate the Markov correction, the range of $x^{(0)}(k)$ (or of its relative errors) is partitioned into m states

$$S_1, S_2, \ldots, S_m,$$

with intervals $[L_i, U_i)$ and representative (midpoint) values

$$c_i = \frac{L_i + U_i}{2}, \quad i = 1, \ldots, m.$$

The state at time k is defined by

$$s_k = i \quad \text{if } x^{(0)}(k) \in [L_i, U_i).$$

From the observed state sequence $s_1, \ldots, s_n$ we obtain transition counts

$$n_{ij} = \#\{k : s_k = i,\ s_{k+1} = j\},$$

and the empirical transition probabilities

$$p_{ij} = P(s_{k+1} = j \mid s_k = i) = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}}, \quad i = 1, \ldots, m.$$

Let $P = [p_{ij}]_{m\times m}$ denote the transition matrix.

Assume that at time j the current state is $s_j = i$. The one–step Grey–Markov forecast is then taken as the conditional expectation of the next observation given the current state:

$$\tilde{x}^{(0)}(j+1) = E\big[x^{(0)}(j+1) \mid s_j = i\big] \approx \sum_{k=1}^{m} p_{ik}\, c_k.$$

Equivalently, if the states are defined on relative deviations δ_k (e.g. midpoints of relative–error intervals), the Grey–Markov forecast can be written as a Markov correction of the Grey forecast:

$$\tilde{x}^{(0)}(j+1) = \hat{x}^{(0)}(j+1)\left(1 + \sum_{k=1}^{m} p_{ik}\, \delta_k\right).$$

$$x_j^{(0)}\big(p_j + (a_j + b_j)/2\%\big). \tag{12}$$

3 Numerical simulations/sensitivity analysis

Given a baseline salary scale of CONPSS 1 from the table above, we assume a positive income salary for individuals in the last 4 years [105000, 115000, 125000, 135000] then when combined with the current salary we have that:

$$x^{(0)} = [x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5)] = [105000, 115000, 125000, 135000, 145500]$$

We will compute the Grey Model (GM(1,1)) for the input sequence:

$$X^{(0)} = [105000, 115000, 125000, 135000, 145500]$$

Compute the Accumulated Generating Operation (AGO) sequence $X^{(1)}$:

$$X^{(1)}(k) = \sum_{i=1}^{k} X^{(0)}(i)$$

$$X^{(1)} = [105000, 220000, 345000, 480000, 625500]$$

$$B = \begin{bmatrix} -0.5(105000 + 220000) & 1 \\ -0.5(220000 + 345000) & 1 \\ -0.5(345000 + 480000) & 1 \\ -0.5(480000 + 625500) & 1 \end{bmatrix} = \begin{bmatrix} -162500 & 1 \\ -282500 & 1 \\ -412500 & 1 \\ -552750 & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ X^{(0)}(4) \\ X^{(0)}(5) \end{bmatrix} = \begin{bmatrix} 115000 \\ 125000 \\ 135000 \\ 145500 \end{bmatrix}$$

Using the least squares method, solve for a and b:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y$$

Compute $B^T B$

$$B^T B = \begin{bmatrix} -162500 & -282500 & -412500 & -552750 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -162500 & 1 \\ -282500 & 1 \\ -412500 & 1 \\ -552750 & 1 \end{bmatrix}$$

$$B^T B = \begin{bmatrix} 162500^2 + 282500^2 + 412500^2 + 552750^2 & -162500 - 282500 - 412500 - 552750 \\ -162500 - 282500 - 412500 - 552750 & 4 \end{bmatrix}$$

$$B^T B = \begin{bmatrix} 581901312500 & -1410250 \\ -1410250 & 4 \end{bmatrix}$$

Compute $B^T Y$

$$B^T Y = \begin{bmatrix} -162500 \cdot 115000 - 282500 \cdot 125000 - 412500 \cdot 135000 - 552750 \cdot 145500 \\ 115000 + 125000 + 135000 + 145500 \end{bmatrix}$$

$$B^T Y = \begin{bmatrix} -1.87 \times 10^{10} - 3.53 \times 10^{10} - 5.57 \times 10^{10} - 8.04 \times 10^{10} \\ 520500 \end{bmatrix}$$

$$B^T Y = \begin{bmatrix} -190112625000 \\ 520500 \end{bmatrix}$$

We need to solve:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y$$

We are tasked with solving the system of equations:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y$$

Given:

$$B^T B = \begin{bmatrix} 581901312500 & -1410250 \\ -1410250 & 4 \end{bmatrix}, \quad B^T Y = \begin{bmatrix} -190112625000 \\ 520500 \end{bmatrix}$$

$$\frac{1}{mq - np} \begin{bmatrix} q & -n \\ -p & m \end{bmatrix}$$

For B^TB:

$$m = 581901312500, \quad n = -1410250, \quad p = -1410250, \quad q = 4$$

The determinant of B^TB is:

$$\det(B^TB) = mq - np = (581901312500)(4) - (-1410250)(-1410250)$$

$$\det(B^TB) = 2327605250000 - 1988805062500 = 338800187500$$

The inverse of B^TB is:

$$(B^TB)^{-1} = \frac{1}{338800187500}\begin{bmatrix} 4 & 1410250 \\ 1410250 & 581901312500 \end{bmatrix}$$

compute:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^TB)^{-1}B^TY$$

Substitute $(B^TB)^{-1}$ and B^TY:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{338800187500}\begin{bmatrix} 4 & 1410250 \\ 1410250 & 581901312500 \end{bmatrix}\begin{bmatrix} -190112625000 \\ 520500 \end{bmatrix}$$

$$a = \frac{1}{338800187500}\left(4 \cdot (-190112625000) + 1410250 \cdot 520500\right)$$
$$a = \frac{1}{338800187500}\left(-760450500000 + 734035125000\right)$$
$$a = \frac{1}{338800187500}\left(-26415375000\right)$$
$$a = -0.07796 \approx -0.08$$

$$b = \frac{1}{338800187500}\left(1410250 \cdot (-190112625000) + 581901312500 \cdot 520500\right)$$
$$b = \frac{1}{338800187500}\left(-268106329406250000 + 302879633156250000\right)$$
$$b = \frac{1}{338800187500}\left(34773303750000000\right)$$
$$b = 102636.6$$

Substitute $a = -0.077$, $b = 102636.6$, and $X^{(0)}(1) = 105000$:
The Grey Model solution is given by:

$$\hat{X}^{(1)}(k+1) = \left(X^{(0)}(0) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}$$

Substituting the given values:

$$\frac{b}{a} = \frac{102636.6}{-0.077} = -1332942.85$$

$$X^{(0)}(0) - \frac{b}{a} = 105000 + 1332942.85 = 1437942.85$$

Thus, the final equation becomes:

$$\hat{X}^{(1)}(k+1) = 1437942.85e^{0.077k} - 1332942.85$$

Table 3: Comparison of Actual and Predicted Values with Relative Error

k	Actual $X_0(k)$	Predicted $X_0(k)$	Relative Error (%)
1	105000	105000.00	0.00
2	115000	115095.93	+0.08
3	125000	124308.45	-0.55
4	135000	134258.36	-0.55
5	145500	145004.67	-0.34
6	155500	156611.15	+0.71
7	165500	169146.63	+2.20
8	175500	182685.48	+4.09
9	185500	197308.00	+6.37
10	195500	213100.95	+9.00
11	205500	230157.99	+11.99

According to relative error (see Table),five states are divided as follows, and the circumstances that relative error lies in the state are listed

- **State 1:** $(-1, 0]$
- **State 2:** $(0, 1]$
- **State 3:** $(1, 2]$
- **State 4:** $(2, 4]$
- **State 5:** $(4, 12)$

Table 4: Comparison of Actual and Predicted Values with Relative Error States for the first five input

k	Actual $X_0(k)$	Predicted $X_0(k)$	Relative Error (%)	State
1	105000	105000.00	0.00	1
2	115000	115095.93	+0.08	2
3	125000	124308.45	-0.55	1
4	135000	134258.36	-0.55	1
5	145500	145004.67	-0.34	1

3.1 Transition matrix P for individual with certificate

The transition matrix P of individuals with certificates is given by:

$$P = \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

As shown in Table 4, we apply the transision probability to the initia state and obtain

The given data shows a system transitioning through different states in a fixed, sequential pattern over five years, moving from State 1 to State 5 without any reversals or skips. Since each transition is

Year	Initial	Transition	State 1	State 2	State 3	State 4	State 5
1	1	1	1	0	0	0	0
2	2	2	0	1	0	0	0
3	1	3	0	0	1	0	0
4	1	4	0	0	0	1	0
5	1	5	0	0	0	0	1
Sum			1	1	1	1	1

deterministic meaning the system always moves to the next state with 100% certainty and never returns to previous states which is the transition probabilities between non-consecutive states are effectively zero. In such cases, the Markov chain component of the Grey-Markov model fails to provide additional predictive insight because there is no variability or randomness in the state transitions. As a result, the Grey-Markov model simplifies to the basic Grey Model (GM(1,1)), which forecasts values based solely on the original data trend without incorporating transition probabilities. This occurs because the strict, unchanging progression of states means the Markov adjustment has no meaningful impact on the forecast, making the Grey-Markov output identical to that of the simpler Grey Model.

Table 5: Comparison of Real and Forecasted Values with State Transition

Year	Real Value	Grey Forecast	Grey Relative Error (%)	State	Grey Markov Forecast	Grey Markov Relative Error (%)
1	115000	115095.93	0.08	1	115095.93	0.08
2	125000	124308.45	-0.55	2	124308.45	-0.55
3	135000	134258.36	-0.55	1	134258.36	-0.55
4	145500	145004.67	-0.34	1	145004.67	-0.34
5	155500	156611.15	0.71	1	156611.15	0.71

3.2 Transition matrix P for individual without educational certificate

The transition matrix P of individuals without educational certificates is given by

$$P = \begin{bmatrix} 0.356 & 0.4 & 0.16 & 0.064 & 0.02 \\ 0 & 0.14 & 0.4 & 0.16 & 0.30 \\ 0 & 0 & 0.24 & 0.4 & 0.36 \\ 0 & 0 & 0 & 0.7 & 0.3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The Grey-Markov forecasting model [12] is obtained by forecasting value of original data sequence, which is obtained according to:

$$y_j = x(0)_j \left(p_j + \frac{a_j + b_j}{2} \% \right) for \quad j = 1, 2, 3, 4, 5$$

$$115095.93 \times \left(1 + \frac{1}{200} \right) = 115671.41$$

Year	Initial	Transition	State 1	State 2	State 3	State 4	State 5
1	1	1	0.356	0	0	0	0
2	2	2	0.4	0.14	0	0	0
3	1	3	0.16	0.4	0.24	0	0
4	1	4	0.064	0.16	0.4	0.7	0
5	1	5	0.02	0.3	0.36	0.3	1
Sum			1	1	1	1	1

$$124308.45 \times \left(1 - \frac{1}{200}\right) = 123686.91$$

$$134258.36 \times \left(1 - \frac{1}{200}\right) = 134929.65$$

$$145004.67 \times \left(1 - \frac{1}{200}\right) = 145729.69$$

$$156611.15 \times \left(1 + \frac{1}{200}\right) = 157394.21$$

Table 6: Comparison of Real and Forecasted Values with State Transition

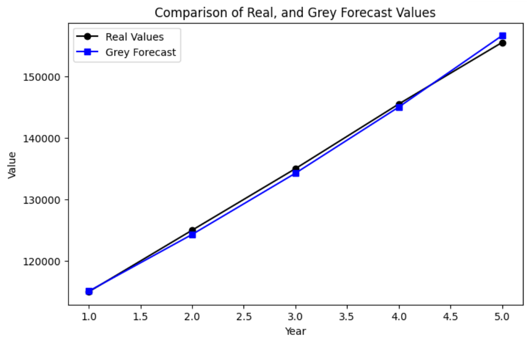
Year	Real Value	Grey Forecast	Grey Relative Error (%)	State	Grey Markov Forecast	Grey Markov Relative Error (%)
1	115000	115095.93	+0.08	2	115671.41	+0.58
2	125000	124308.45	-0.55	1	124929.99	-0.06
3	135000	134258.36	-0.55	1	134929.65	-0.05
4	145500	145004.67	-0.34	1	145729.69	+0.16
5	155500	156611.15	+0.71	2	157394.21	+1.22

3.3 Result analysis: salary progression and tax reform implications

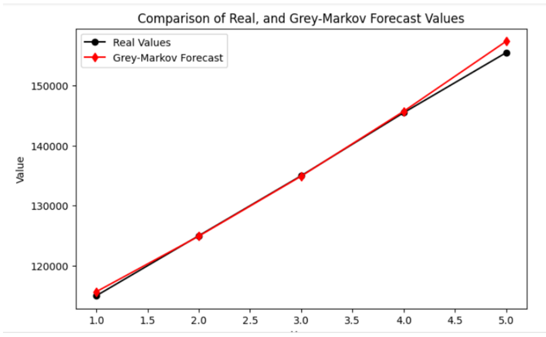
In light of ongoing discussions surrounding tax reform and income equity, a forecasting experiment was conducted using historical salary progression data on the CONPSS 1 scale. This scale represents the baseline earnings for civil servants or entry-level public sector employees in many developing economies. The objective was to examine income growth trends and assess how tax policy adjustments might influence fiscal planning, wage structure, and individual take-home pay.

The model employed the Grey Model GM(1,1) integrated with a Markov chain transition matrix, which enabled the analysis of the non-linear growth of salaries over a period of five years, including projections up to six additional years. The initial income sequence:

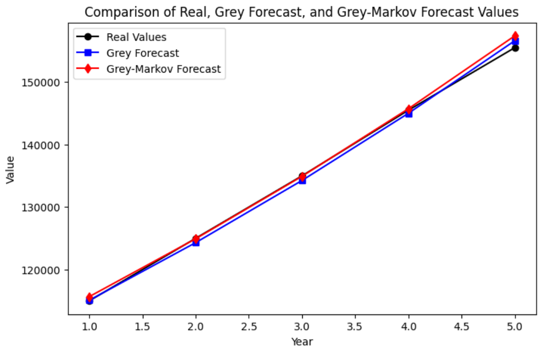
$$\{105,000; 115,000; 125,000; 135,000; 145,500\}$$



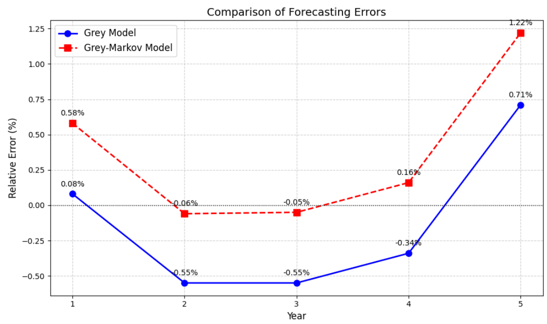
(a) Real vs Grey Forecast Values



(b) Real vs Grey-Markov Forecast Values



(c) Real, Grey, and Grey-Markov Forecast Values



(d) Comparison of Forecast Errors

Figure 6: Comparative Analysis of Forecasting Methods are shown in figure (a-d)

revealed a consistent upward trend, mirroring annual increments tied to either inflation adjustments or policy-driven raises.

Through the Accumulated Generating Operation (AGO), a smoothed sequence was derived to reduce volatility and better understand underlying trends. The model coefficients:

$$a = -0.08, \quad b \approx 102,636.6$$

indicated a moderately exponential growth pattern. Predicted values remained close to actual salaries, with relative errors largely falling within $\pm 1\%$ for the first five years, demonstrating high predictive reliability.

The transition states classified these deviations into **five defined bands** (State 1 to State 5), reflecting the degree of divergence from actuals. Early years remained in States 1 and 2, indicating low deviation, while later years shifted to higher states (3 to 5), showing greater uncertainty and variability in income forecasting.

Tax reform implications

The simulation has direct implications for tax policy and reform strategies:

1. **Progressive Tax Bracket Adjustments:** As salaries increase predictably, tax brackets must be adjusted periodically to prevent *fiscal drag*, where workers unintentionally shift into higher tax bands due to inflation-driven increments rather than actual wealth growth.

2. **Revenue Forecasting:** Reliable salary progression models like the Grey-Markov model support realistic revenue projections. By anticipating income levels, the government can better estimate PAYE (Pay-As-You-Earn) contributions and adjust tax policies accordingly.

3. **Equity Considerations:** The consistent growth pattern suggests that while entry-level incomes are increasing, disparities may emerge if higher salary scales grow exponentially faster. Reforms should aim to flatten regressive tax influences, ensuring equity across income groups.

4. **Policy Buffering for Inflation:** With forecasted salary increases up to 230,157.99 in year 11, aligning tax thresholds with projected salary growth helps preserve real income and support purchasing power stability.

5. **Behavioral Adaptation and State Transitions:** The transition matrix reveals that income categories tend to evolve into higher error states over time. This highlights the dynamic and uncertain nature of salary forecasting under external shocks, warranting flexible and adaptive tax frameworks.

Conclusion

The integration of Grey and Markov models in salary forecasting proves to be a robust approach for analyzing wage dynamics. When contextualized within a tax reform framework, these results underscore the importance of responsive and data-driven taxation systems. As economies strive for fairness and efficiency, leveraging such models ensures that tax reforms are grounded in empirical insight, enhancing both policy effectiveness and public trust.

Competing Interests

The author(s) declare that there are no financial, personal, or professional competing interests that could have influenced the research reported in this work.

References

[1] Hamsici, O. C. and Martinez, A. M. (2014). Multiple ordinal regression by maximizing the sum of margins. IEEE Transactions on Neural Networks and Learning Systems, 25(12), 2262–2275.

[2] Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2013). Bacterial persistence as a phenotypic switch. PLOS ONE, 8(6), e62814.

[3] Sasu, D. D. (2023). Employment in Nigeria - Statistics & Facts. Statista. https://www.statista.com/topics/7055/employment-in-nigeria/ [Accessed: Day Month Year].

[4] Punch Newspaper. (2023). 80.4% of Nigerian employment in informal sector, says World Bank. https://punchng.com/80-4-of-nigerian-employment-in-informal-sector-says-wbank/ [Accessed: Day Month Year].

[5] Ekugo, N. (2024). 10 key takeaways from Nigeria's 2024 labour force survey. *BusinessDay*. https://businessday.ng/news/article/10-key-takeaways-from-nigerias-2024-labour-force-survey/ [Accessed: Day Month Year].

[6] Li, X., and Xing, Q. (2021). Prediction of income of rural residents based on Markov model under the background of rural revitalization. Journal of Rural Studies, 82, 1–10.

[7] Shodiq, M., Warsito, B. and Gernowo, R. (2018). The implementation of grey forecasting model for forecast of captured fisheries production. Jurnal Ilmu Komputer dan Informasi, 9(4), 195–202.

[8] Mao, Z. and Sun, J. (2011). Application of Grey-Markov model in forecasting fire accidents. Procedia Engineering, 11, 314–318.

[9] Aminu, A. (2019). Characterising graduate unemployment in Nigeria as education-job mismatch problem. African Journal of Economic Review, 7(2), 113–130.

[10] Edem, I.J., Adebiyi, K.A., and Ayo, C.K. (2015). A modified grey-Markov fire accident model based on information turbulence indices and restricted residuals. International Journal of Management Science and Engineering Management, 10(3), 191–199.

[11] Seber, G.A.F. and Lee, A.J. (2003). Linear regression analysis (2nd ed.). John Wiley & Sons.

[12] Tribune Online. (2025). Minimum wage: Breakdown of salary structure for civil servants. https://tribuneonlineng.com/minimum-wage-breakdown-of-salary-structure-for-civil-servants/ [Accessed: Day Month Year].

[13] Nguyen, P., Tsai, J., Kayral, I.E., and Lin, M. (2021). Unemployment rates forecasting with grey-based models in the post-COVID-19 period: A case study from Vietnam. Sustainability, 13(14), 7879.

[14] Lawal, A., Didigwu, N.E., Saidu, D.Y., Sadiq, S.L., and Khadeejah, J.A. (2021). Application of Grey-Markov model for forecasting Nigeria annual rice production. Scientia Africana, 20(3), 73–84.

[15] Huda, N. and Imro'ah, N. (2023). Grey Markov (1,1) model for forecasting the percentage of the population that experienced health complaints in Indonesia. Jurnal Matematika UNAND, 12(2), 108–120.

[16] Obia, S.O. and Akeyede, I. (2022). Forecasting natural gas consumption in Nigeria using the modified grey model (MGM(1,1, b)). African Scientific Reports, 1(2), 115–122.

[17] Zhang, H. and Chen, Y. (2021). Analysis and application of grey-Markov chain model in tax forecasting. Journal of Mathematics, 2021, 9918411.

[18] D'Amico, G., Karagrigoriou, A., and Vigna, V. (2023). Forecasting the power generation mix in Italy based on grey Markov models. European Journal of Operational Research, 307(1), 399–412.

[19] Hu, Y.-C., Wu, G., and Jiang, P. (2023). [Title of the article]. Journal of Hospitality & Tourism Research, 47(5), 775–799.

[20] Atoyebi, O. M. (2024). The Nigerian Tax Bill 2024 at a glance: Highlights and key provisions. Omaplex. https://omaplex.com.ng/the-nigerian-tax-bill-2024-at-a-glance-highlights-and-key-provisions/ [Accessed: Day Month Year].