IRAQI
Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2024) 14 (4): 387-401

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN 2222-2995
University of Kirkuk Journal For Administrative and Economic Science

Hussein Mohammad Mahmood Faqe. Estimate the Parameters of the Bootstrap Grey Model GM (1,1) and use them in the Hybrid Grey Markov Chain Model. *University of Kirkuk Journal For Administrative and Economic Science* **(2024) 14 (4):387-401.**

Estimate the Parameters of the Bootstrap Grey Model GM (1,1) and use them in the Hybrid Grey Markov Chain Model

Mohammad Mahmood Faqe Hussein [1]

[1] *University of Sulaimani, College of Administration and Economics, Sulaimani, Iraq*

Mohammad.faqe@univsul.edu.iq[1]

Abstract: This work investigates the estimation of Bootstrap Grey Model GM (1,1) parameters and their use in a Hybrid Grey Markov Chain Model to improve forecasting precision. The Grey Model GM (1,1) works effectively with dynamic systems that have little or no information and tiny datasets. The Accumulated Generating Operation (AGO) is used to modify the original time series data, and a data matrix is created in order to estimate the model's parameters, a and b, using least squares estimation. In order to improve the robustness of the parameter estimations, bootstrap resampling is used, which yields numerous resampled datasets from which the parameters' mean and confidence ranges are obtained. These improved parameters are then integrated into the Hybrid Grey Markov Chain Model, which combines the advantages of Markov chains and grey system theory. In order to modify the forecasts based on probabilistic state transitions taken from the World Health Organization's Global Tuberculosis Report, this hybrid model first applies the grey model for starting state forecasting. Markov chains are then used. According to the results, the number of persons affected with this sort of sickness starts to decline in the upcoming years when we use the Hybrid Grey Markov Chain Model and Estimate the Parameters of the Bootstrap Grey Model GM (1,1) to get accurate and effective results.

Keywords: Accumulated Generating Operation (AGO), Bootstrap Grey Model, Hybrid Grey Markov Chain Model, and Probabilistic State Transitions

تقدير معاملات نموذج بوتستراب الرمادي GM (١،١) واستخدامها في نموذج سلسلة ماركوف الرمادية الهجينة

أ.م.د. محمد محمود فقي حسين [١]

[١] *جامعة السليمانية – كلية الادارة والاقتصاد، السليمانية، العراق*

المستخلص: يستكشف هذا البحث تقدير معلمات نموذج بوتستراب الرمادي GM (١،١) وتطبيقها ضمن نموذج سلسلة ماركوف الرمادية الهجينة لتحسين دقة التنبؤ. يعد نموذج الرمادي GM (١،١) مناسبًا تمامًا لمجموعات البيانات الصغيرة والأنظمة الديناميكية ذات المعلومات غير المؤكدة وغير المكتملة. تتضمن المنهجية تحويل بيانات السلسلة الزمنية الأصلية من خلال عملية التوليد المتراكمة (AGO) وإنشاء مصفوفة بيانات لتقدير معلمات النموذج

a و b باستخدام تقدير المربعات الصغرى. لتعزيز قوة تقديرات المعلمات، يتم استخدام إعادة أخذ العينات من بوتستراب، مما يوفر مجموعات بيانات متعددة تم إعادة أخذ عينات منها والتي يتم اشتقاق متوسط وفترات الثقة للمعلمات منها. بعد ذلك، يتم دمج هذه المعلمات المكررة في نموذج سلسلة ماركوف الرمادية الهجينة، والذي يعمل على تآزر نقاط القوة في نظرية النظام الرمادي وسلاسل ماركوف. يستخدم هذا النموذج الهجين أولاً النموذج الرمادي للتنبؤ بالحالة الأولية، يليه تطبيق سلاسل ماركوف لضبط التوقعات بناءً على انتقالات الحالة الاحتمالية المستمدة من منظمة الصحة العالمية، تقرير مكافحة السل في العالم. تظهر النتائج أن استخدام تقدير معلمات نموذج بوتستراب الرمادي (١،١) GM واستخدام نموذج سلسلة ماركوف الرمادية الهجين، يعطينا نتائج دقيقة وفعالة، ويبدأ عدد الأشخاص المصابين بهذا النوع من الأمراض في الانخفاض في السنوات القادمة.

الكلمات المفتاحية: نموذج بوتستراب الرمادي، نموذج سلسلة ماركوف الرمادية الهجينة، عملية التوليد المتراكمة (AGO)، مصفوفة انتقال الحالة، انتقالات الحالة الاحتمالية.

Corresponding Author: E-mail: Mohammad.faqe@univsul.edu.iq

Introduction

This research explores the estimation of parameters for the Bootstrap Grey Model GM (1,1), a versatile tool for forecasting in systems with limited and uncertain data. By leveraging the bootstrap method, the model's reliability and precision are significantly enhanced. These optimized parameters are then applied within the Hybrid Grey Markov Chain Model, which synergizes Grey Model predictions with the probabilistic insights of Markov chains. This hybrid framework aims to improve forecasting accuracy by addressing both deterministic and stochastic system behaviors. The study provides a robust methodology for applications in dynamic and complex systems analysis

1. Research objective

A. To develop and apply a reliable resampling technique for parameter estimation for the Bootstrap Grey Model GM (1,1). This entails taking into account data variability, enhancing the robustness of the model, and refining the estimation procedure to increase the accuracy and reliability of the parameter estimates.

B. To improve forecasting accuracy and predictive performance, a hybrid grey Markov chain model will be constructed using the parameters derived from the Bootstrap Grey Model GM(1,1). This entails utilizing the Grey Model's parameters to refine the transition probabilities of the Markov Chain model, resulting in a more accurate and dependable forecasting tool.

2. Research Problem

A well-liked forecasting and time series analysis tool, the GM (1,1) is renowned for its ease of use and efficiency when working with systems that have little data. However, the inherent uncertainties and variability in the data may not be adequately addressed by conventional methods for estimating its characteristics, which could result in forecasts that are less accurate. Because bootstrap procedures provide more robust statistical conclusions and better account for this fluctuation, they may enhance parameter estimation. This study aims to enhance the Bootstrap Grey Model GM (1,1)'s parameter estimation procedure and investigate how it might be integrated with the Hybrid Grey Markov Chain Model to increase forecasting precision and dependability.

3. Literature Review

In many fields, including environmental science, engineering, and economics, time series forecasting is essential. Small, ambiguous, or partial datasets frequently provide difficulties for traditional forecasting methods. Although it might have trouble with parameter estimation, the GM (1,1) is expressly made to deal with these problems. This study of the literature looks at how bootstrap resampling and its incorporation into a hybrid grey Markov chain model have improved parameter estimation for the GM (1,1) model. Chen and Hsu (2006) improved parameter estimation in the GM (1,1) by using bootstrap resampling. According to Zhang et al. (2019), bootstrap techniques can greatly increase the gray model parameters' correctness and stability. A hybrid model that blends Markov chains with gray systems was presented by Liu and Lin (2006). Zhou et al. (2020) demonstrated the efficacy of the hybrid model in forecasting environmental data, whereas Yu et al. (2014) used this hybrid approach to improve forecasting accuracy in stock markets.

4. Methodology

Grey System Theory's foundational GM (1,1) is intended for predicting in situations with sparse, ambiguous, or insufficient data. Developed by Deng Julong in the 1980s, this model predicts the behavior of a single-variable time series using a first-order differential equation. An Accumulated Generating Operation (AGO) is used at the start of the process to change the original dataset, eliminate noise, and highlight the underlying trend. The model's parameters are estimated using least squares or comparable techniques to create the differential equation that captures the dynamics of the system. To return the forecasts to the original data scale, an inverse transformation is lastly used. The GM (1,1) model demonstrates significant efficacy when applied to small datasets, characterized by its simplicity, dependability, and straightforward interpretability. Its versatility allows for a wide range of applications across various disciplines, including economics, engineering, environmental science, and healthcare, thereby establishing it as a valuable forecasting instrument for both practical applications and academic research.

A. Fundamental of Grey Theory

The grey prediction methodology is based on the GM (n, m) model, where n signifies the order of the grey difference equation and m indicates the number of variables involved. A significant portion of prior research in the domain of grey prediction models has focused on the GM (1, 1) model. This particular model effectively balances simplicity with accuracy in its results.[10][11]A sequence of raw data that is non-negative is presented as follows: For a system comprising n variables and differential equations of order mmm, the GM(n,m) model is organized in the manner outlined below.:

$$\frac{d^m w_i^{(1)}(t)}{dt^m} + \sum_{j=1}^{n} a_{ij} w_j^{(1)}(t) = b_i, \quad i = 1,2,\dots,n$$

where:

$w_i^{(1)}(t)$: is the accumulated generating operation (AGO) of the i-th variable over time t,

$\frac{d^m w_i^{(1)}(t)}{dt^m}$: represents the m-th order derivative of $w_i^{(1)}(t)$

a_{ij}: are coefficients that reflect the influence of the j-th variable on the i-th equation,

b_i : represents an external input or grey action affecting $x_i^{(1)}(t)$

$$w_{(i)}^{(0)} = \left\{w_{(1)}^{(0)}, w_{(2)}^{(0)}, \dots, w_{(n)}^{(0)}\right\} \quad , n \geq 4 \qquad (1)$$

Where, n is the sample size.

The (AGO) is employed to smooth out the randomness in the original sequence. By applying AGO, the initial sequence is transformed into a monotonically increasing sequence. The resulting sequence, denoted as $W^{(1)}$, is created by AGO as follows:[2][4][5]

$$w_{(i)}^{(1)} = \{w_{(1)}^{(1)}, w_{(2)}^{(1)}, \dots, w_{(n)}^{(1)}\} \quad , n \geq 4 , \qquad (2)$$

$n \geq 4$, the model can avoid oversimplification, particularly in cases where higher complexity is required. Models with

Where,

$$w_{(1)}^{(1)} = w_{(1)}^{(0)}$$

$$w_{(k)}^{(1)} = \sum_{j=1}^{k} w_{(k)}^{(0)}, \qquad (j = 1,2,\dots,k). \qquad (3)$$

The GM (1, 1) model is formulated by creating a 1-order model for $w_{(k)}^{(1)}$, which is expressed as:

$$\frac{dw_{(k)}^{(1)}}{dk} + aw_{(k)}^{(1)} = u \qquad (4)$$

In this context, the parameters **a** and **u** are referred to as the development coefficient and grey input, respectively, for the above equation we obtain the value of $\widehat{w}_{(k+1)}^{(1)}$:

$$\widehat{w}_{(k+1)}^{(1)} = \left[w_{(1)}^{(0)} - \frac{u}{a}\right] e^{-ak} + \frac{u}{a} \qquad (5)$$

The coefficients $[a,u]^T$ can be obtained by the OLS method:

$$\hat{a} = [a, u]^T = (\beta^T\beta)^{-1}\beta^T Y \qquad (6)$$

In that:

$$Y_n = [w_{(2)}^{(0)}, w_{(3)}^{(0)}, \dots, w_{(n)}^{(0)}]^T$$

$$\beta = \begin{bmatrix} -\frac{1}{2}(w_{(1)}^{(1)} + w_{(2)}^{(1)}) & 1 \\ -\frac{1}{2}(w_{(2)}^{(1)} + w_{(3)}^{(1)}) & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ -\frac{1}{2}(w_{(n-1)}^{(1)} + w_{(n)}^{(1)}) & 1 \end{bmatrix} \qquad (7)$$

In a Grey Model β often represents a matrix (or vector) of coefficients

IAGO is used to find forecast values of rude sequence:

$$\widehat{w}_{(k+1)}^{(0)} = [w_{(1)}^{(0)} - \frac{u}{a}]e^{-ak} + \frac{u}{a} \qquad (8)$$

Therefore, the fitted and forecasted sequence $\widehat{x}^{(0)}$ is given as:

$$\widehat{w}^{(0)} = \left(\widehat{w}_{(1)}^{(0)}, \widehat{w}_{(2)}^{(0)}, \dots, \widehat{w}_{(n)}^{(0)}, \dots\right) \qquad (9)$$

and

$$\widehat{w}_{(1)}^{(0)} = w_{(1)}^{(0)}$$

Where

$\widehat{w}_{(1)}^{(0)}, \widehat{w}_{(2)}^{(0)}, \dots, \widehat{w}_{(n)}^{(0)}$ are called the fitted of GM (1, 1)

while:

$\widehat{w}_{(n+1)}^{(0)}, \widehat{w}_{(n+2)}^{(0)}, \dots, \widehat{w}_{(n+k)}^{(0)}$ are called the forecast values of GM (1, 1).

B. Model Assessment and accuracy:

The test is divided into three parts.[3]

<u>**Residual Test**</u>**:**

<u>**Step1:**</u> The size of the model value and the actual value of the residual point by point inspection is according to the first model to calculate

$w_{(n+1)}^{(0)}$, will be $w_{(i+1)}^{(0)}$ regressive generation $w_{(i)}^{(0)}$, finally calculated the original sequence $w_{(i)}^{(0)}$ and absolute residual sequence and relative residual sequence

$$\delta_{(i)}^{(0)} = \left|w_{(i)}^{(0)} - \widehat{w}_{(i)}^{(0)}\right| \qquad (10)$$

$$\emptyset_{(i)} = \left[\frac{\delta_{(i)}^{(0)}}{w_{(i)}^{(0)}}\right] \times 100 \qquad i = 1,2, \dots \dots, n$$

<u>**Step2:**</u> Calculate the two –step of the relative error as follows:

$$\delta_{min} = min\left|\delta_{(i)}^{(0)}\right| \qquad (11)$$

$$\delta_{max} = max\left|\delta_{(i)}^{(0)}\right| \qquad (12)$$

<u>**Step3:**</u> Compute coefficients of grey incidence

$$\gamma\left(\widehat{w}_{(k)}^{(0)}, w_{(k)}^{(0)}\right) = \frac{\delta_{min} + p.\delta_{max}}{\delta_{oi}(k) + p.\delta_{max}} \qquad (13)$$

<u>**Step4:**</u> Compute the degree of grey incidence

$$\gamma\left(\widehat{w}_{(i)}^{(0)}, w_{(k)}^{(0)}\right)=\frac{1}{n}\sum_{i=1}^{n}\gamma\left(\widehat{w}_{(i)}^{(0)}, w_{(k)}^{(0)}\right) \qquad (14)$$

C. Mean Absolute Percentage Error

The overall accuracy of this model can be measured by its MAPE and it is defined as follows:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{k=2}^{n}\left|\frac{w_{(k)}^{(0)}-\widehat{w}_{(k)}^{(0)}}{w(k)}\right| \times 100 \quad , k = 2,3,..,n \qquad (15)$$

Where $w_{(k)}^{(0)}$ is actual value, $\widehat{w}_{(k)}^{(0)}$ is the forecasted value, and n is the sample size.

Precision Rate

Measures the level of the closeness of the statement of prediction and the actual value, P is calculated as follows:

$$P = 1 - \frac{1}{n}\sum_{k=2}^{n}\left|\frac{w_{(k)}^{(0)}-\widehat{w}_{(k)}^{(0)}}{w(k)}\right| * 100 \quad , k = 2,3,..,n \qquad (16)$$

Table (1): Shows the levels of precision rate

Precision rank	Precision rate
Highly accurate	$P \geq 0.99$
Good	$P \geq 0.95$
Reasonable	$P \geq 0.90$
Inaccurate	$P \leq 0.90$

Posterior Difference Test

Posterior difference inspection is i.e. on residual distribution statistical properties of the inspection.

(1) Calculate the mean of the original sequence:

$$\overline{w}^{(0)} = \frac{\sum_{i=1}^{n} w_{(i)}^{(0)}}{n}$$

(2) Calculating the original sequence $x^{(0)}$ mean square error :

$$S_1 = \sqrt{\frac{\sum_{i=1}^{n}\left[w_{(i)}^{(0)} - \overline{w}^{(0)}\right]^2}{n-1}}$$

(3) Calculating residual mean:

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\bar{\delta}_{(i)}^{(0)}$$

(4)

$$S_2 = \sqrt{\frac{\sum_{i=0}^{n}\left[\delta_{(i)}^{(0)} - \bar{\delta}\right]^2}{n-1}}$$

(5) Calculation variance ratio C:

$$c = S_2/S_1 \qquad (17)$$

(6) Calculation small residual probability:

$$P = P\left\{\left|\delta_{(i)}^{(0)} - \bar{\delta}\right| < S_0 = 0.6745 S_1\right\}$$

Make

$$e_i = \left|\delta_{(i)}^{(0)} - \bar{\delta}\right|$$

Table (2): Show the posterior difference test discriminant reference table

P	C	Model accuracy
> 0.95	< 0.35	Optimal
> 0.80	< 0.5	Qualified
> 0.70	> 0.65	Forced to pass the exam
< 0.70	> 0.65	The unqualified

D. The bootstrap in linear regression

The bootstrap method is a statistical approach employed to estimate population parameters by calculating averages from multiple small data samples. This technique is essential as it entails selecting observations from a large dataset individually, with each observation being returned to the dataset after selection.

(1) Bootstrapping in linear regression

It is important to recognize that bootstrapping serves not only as a useful tool for estimating standard errors but also for creating confidence intervals and performing hypothesis tests. Consequently, it is advisable to employ bootstrapping techniques when working with data that may not conform to the assumptions required by conventional methods.

(2) Bootstrapping technique

The bootstrapping method is a technique used to estimate statistics about a population by taking repeated samples from a dataset, allowing for the same data points to be chosen more than once. Here's a breakdown of how it works, along with the formula:

(3) Steps in bootstrapping

(a) Begin with an initial sample of size n.
(b) **Resampling**: Draw new samples of size n from the original sample, allowing for replacement.
(c) **Statistic Calculation**: For each resample, calculate the relevant statistic (like variance, mean, or median) and create a distribution based on these values.
(d) **Estimate and Confidence Intervals**: Use the distribution of bootstrapped statistics to estimate the population parameter and its confidence intervals.

(4) Formulation

Starting with an original sample $\omega = (\omega_1, \omega_2, \dots, \omega_n)$

Resamples: Generate β resamples, each containing n elements. Let Z_b^* denoted the $b-th$ resample $(b = 1,2, \dots, \beta)$

$$\omega_b^* = (\omega_{b1}^*, \omega_{b2}^*, \dots, \omega_{bn}^*) \qquad (18)$$

where each W_{bi}^* is selected with replacement from the original sample.

(a) Statistic Calculation: For each resample, compute the statistic $\widehat{\emptyset}$. For instance, if the statistic is the mean: $\widehat{\emptyset}_b^* = \frac{1}{n}\sum_{i=1}^{n} \omega_{bi}^*$ (19)

(b) Bootstrap Distribution: The set of bootstrapped statistics $\{\widehat{\emptyset}_1^* *, \widehat{\emptyset}_2^*, \dots, \widehat{\emptyset}_B^*\}$ creates the bootstrap distribution.

(c) Estimation: To estimate the population parameter $\emptyset$, take the average of the bootstrap statistics:

$$\hat{\theta}_{\text{bootstrap}} = \frac{1}{\beta}\sum_{i=b}^{\beta} \hat{\emptyset}_b^* \qquad (20)$$

(d) Confidence Intervals: Determine confidence intervals using percentiles from the bootstrap distribution. For a 95% confidence interval, the 2.5th and 97.5th percentiles of the bootstrap statistics are typically used.

$$\mathrm{CI}_{95\%} = \left(\hat{\emptyset}^*_{\frac{\beta-0.025}{100}}, \hat{\emptyset}^*_{\frac{\beta-0.975}{100}}\right) \qquad (21)$$

E. Building Grey-Markov Chain Model

The remaining errors are first divided into q states, which are defined as S_1, S_2, … , S_q, and each state meets the equiprobability criterion. This is the first stage in structure the GM. The probability from state S_i to state S_j, which yields the transition matrix P, is then used to create the transition matrix. Determine the P_{ij} transition probabilities between the states State $_i$ and State $_j$:

$$P_{ij} = \frac{\#\text{ of transition from State}_i \text{ to State}_j}{\sum \text{State}_i}$$

$$P^{(1)} = \begin{pmatrix} P^{(1)}_{(11)} & P^{(1)}_{(12)} & \cdots & P^{(1)}_{(1q)} \\ P^{(1)}_{(21)} & P^{(1)}_{(22)} & \cdots & P^{(1)}_{(2q)} \\ \vdots & \vdots & \vdots & \vdots \\ P^{(m)}_{(q1)} & P^{(m)}_{(q2)} & \cdot & P^{(m)}_{(qq)} \end{pmatrix}$$

$$P^{(m)} = \begin{pmatrix} P^{(m)}_{(11)} & P^{(m)}_{(12)} & \cdots & P^{(m)}_{(1q)} \\ P^{(m)}_{(21)} & P^{(m)}_{(22)} & \cdots & P^{(m)}_{(2q)} \\ \vdots & \vdots & \vdots & \vdots \\ P^{(m)}_{(q1)} & P^{(m)}_{(q2)} & \cdot & P^{(m)}_{(qq)} \end{pmatrix}$$

Where $P^{(m)}_{ij} = \frac{M^{(m)}_{ij}}{m_i}$, $(i, j = 1,2,3, \dots, L)$, $M^{(m)}_{ij}$ **stand for the transition form** S_i to S_j in m **steps and** m_i **is a number of state** S_i **.**

Next, the remaining error must be confirmed.
Let the interval median in $[S_{i-}, S_{i+}]$ be residual error forecasting value as follows:

$$\hat{e} = \frac{1}{2}[S_{i-} + S_{i+}]$$

Hence, the Grey-Markov model is obtained as:

$$\hat{Y}(k+1) = [1+\hat{e}]\hat{X}^{(0)}(k+1)$$

Where $$\hat{W}^{(0)}(k+1) = \hat{W}^{(1)}(k+1) - \hat{W}^{(1)}(1)$$

Hence

$$\widehat{Y}(k+1) = [1 + \frac{1}{2}((S_{i-} + S_{i+})]\widehat{W}^{(0)}(k+1)$$

5. Result and Discussion

In Iraq, tuberculosis (TB) is a significant public health concern, much like it is globally. Although it can affect other body organs, tuberculosis (TB), which is caused by the bacteria Mycobacterium tuberculosis, mainly affects the lungs. People who have active pulmonary tuberculosis can spread the disease through their speech, sneeze, or cough. Public health campaigns to increase awareness about TB prevention and the importance of following treatment regimens, early diagnosis, and efficient treatment protocols are some of the techniques used in Iraq's TB management efforts. Key players in these initiatives are the World Health Organization (WHO) and regional health authorities, who guarantee correct data reporting and successful intervention execution. This page's statistics came from the World Health Organization's Global Tuberculosis Report. in this page https://data.worldbank.org/indicator/SH.TBS.INCD?locations=IQ

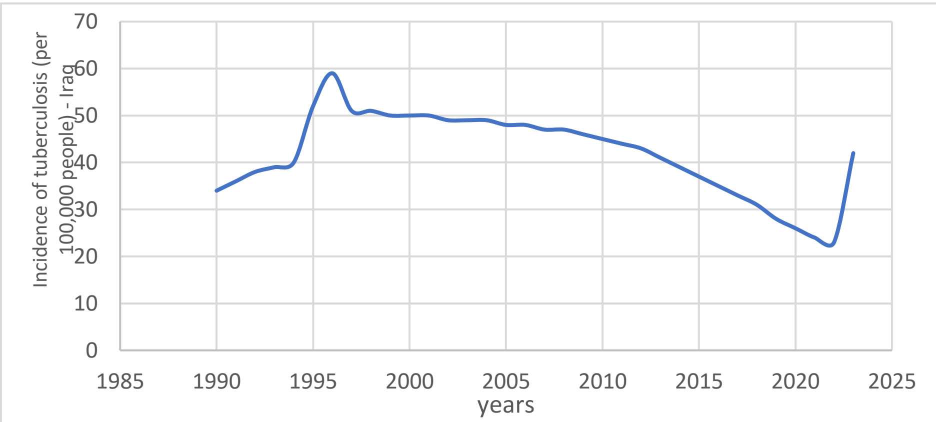

Figure (1): Represent the Prevalence of tuberculosis (per 100,000 people) - Iraq

The first set of statistics for Iraq from 1990 to 2023 is shown in the above graphic. After that, the researchers forecasted the incidence of tuberculosis for the years 1990–2023 using the Bootstrap Grey-Markov. Following the guidelines provided in the theoretical part, the following are the processes to predict the prevalence of tuberculosis:

Building GM (1,1) with and without using Bootstrap forecasting Model

Based on the data of Prevalence of tuberculosis (per 100,000 people) – Iraq of the Iraq from 1990 to 2023.The results are a follow using xlstata package:

Table (3): Explain The Coefficient of the GM(1,1) **without using Bootstrapping**

Coefficients					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	51.851	2.642		19.626	0.000
X	0.012	0.003	0.601	4.187	0.000
a. Dependent Variable: Y					

The equation y=51.851+0.012X describes how y changes as X changes. It shows a linear relationship where y increases by 0.012 units for every one-unit increase in X, starting from a value of 51.851 when X is zero.

Bootstrap Specifications:

Number of Samples= 500 : Confidence Interval Level =95%

Table (4): Explain The Coefficient of the GM(1,1) using **Bootstrapping**

Bootstrap for Coefficients:						
Model	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BC 95% Confidence Interval	
					Lower	Upper
(Constant)	51.851	0.472	3.620	0.002	46.211	60.974
X	0.012	0.000	0.004	0.012	0.006	0.021

Number of Samples= 1000 : Confidence Interval Level =95%

Table (5): Explain The Coefficient of the GM(1,1) using **Bootstrapping**

Bootstrap for Coefficients							
Model		B	Bootstrap				
			Bias	Std. Error	Sig. (2-tailed)	BC 95% Confidence Interval	
						Lower	Upper
1	(Constant)	51.851	0.413	3.600	0.001	45.467	60.173
	X	0.012	0.000	0.004	0.013	0.005	0.022

Number of Samples= 2000 : Confidence Interval Level =95%

Table (6): Explain The Coefficient of the GM(1,1) using **Bootstrapping**

Bootstrap for Coefficients						
Model	B	Bootstrap				
		Bias	Std. Error	Sig. (2-tailed)	BC 95% Confidence Interval	
					Lower	Upper
(Constant)	51.851	0.291	3.650	0.000	46.100	59.773
X	0.012	0.000	0.004	0.018	0.006	0.021

▪$\hat{\beta}_0$=51.851: In the linear regression model, this is the intercept. When all independent variables are zero, it shows the dependent variable's expected value. A bootstrap of the 2000 estimated intercepts yields an average of 51.851.

▪$\hat{\beta}_1$=0.012: The linear regression model's slope in relation to one of the predictors is this. It shows how much the dependent variable changes on average when the predictor changes by one unit. Bootstrapping yields an average of 2000 estimated slopes of 0.012.

▪**Bias for $\hat{\beta}_0$**=0.291: The discrepancy between the initial sample estimate and the bootstrap average is known as the bias. The bootstrap estimates of $\hat{\beta}_0$ are generally marginally higher than the initial estimate, as seen by the bias of 0.291 for $\hat{\beta}_0$

▪**Bias for $\hat{\beta}_1$**=0.000: Since the bias for $\hat{\beta}_1$ is zero, the bootstrap estimates of $\hat{\beta}_1$are consistent with the original estimate on average.

▪**95% Confidence Interval for $\hat{\beta}_0$ (46.100 to 59.773)**: The bootstrap distribution of $\hat{\beta}_0$ is used to compute this interval. This indicates that there is a 95% confidence interval between 46.100 and 59.773 for the real value of the intercept $\hat{\beta}_0$.

▪**95% Confidence Interval for $\hat{\beta}_1$ (0.006 to 0.021)**: The bootstrap distribution of $\hat{\beta}_1$ is used to compute this interval. This indicates that we have a 95% confidence level that the slope $\hat{\beta}_1$ true value falls between 0.006 and 0.021.

▪**Standard Error for $\hat{\beta}_0$**=3.650: This represents the bootstrap estimations' standard deviation. $\hat{\beta}_0$=51.851 and $\hat{\beta}_1$=0.012 are the average bootstrapping estimates for the coefficients. The bootstrapping technique offers a means of evaluating the estimates' variability and confidence

intervals, providing information about the accuracy and dependability of the model parameters. The range that the genuine parameter values most likely fall inside is indicated by the 95% confidence interval, which quantifies the uncertainty surrounding these point estimates.
Since bootstrapping resamples from the same initial dataset, the values of $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ (intercept and slope) stay consistent across sample sizes. The underlying population distribution is about the same for all samples, regardless of size, which leads to consistent estimates of $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ as they approach the true values.

Model Specifications

The **Incidence of tuberculosis** of Iraq from 1990 to 2023 forms and original discrete data sequences:

$\mathbf{w}^{(0)}$=[34,36,38,39,40,52,59,51,51,50,50,50,49,49,49,48,48,47,47,46,45,44,43,41,39,37,35,33,31,28,26,24,23,42]

The following data series can be calculated through the Accumulating Generation Operator (AGO)
$\mathbf{w}^{(1)}$=[7.96,15.70,23.22,30.52,37.60,44.46,51.10,57.52,63.72,70.01,76.37,82.81,89.32,95.92,102.68,109.60,116.68,123.92,131.32,138.67,145.97,153.22,160.42,167.57,174.65,181.66,188.60,195.47,202.27,208.98,215.60,222.13,228.57,234.92,241.22,247.46,253.66,259.80,265.89,271.89,277.80,283.63,289.36,295.01,300.56,306.02,311.39,316.67,321.86,326.96,331.95,336.86,341.67,346.38,351.04,355.62,360.15,364.61,369.01,373.38,377.72,382.03,386.31,390.56,394.69,398.72,402.63,406.42,410.10,413.74,417.33,420.87,424.36,427.81]

Further, data Matrices β and x can be established as:

$$\beta = \begin{bmatrix} -11.8285 & 1 \\ -19.458 & 1 \\ -26.868 & 1 \\ . & . \\ \vdots & \vdots \\ -419.097 & 1 \\ -8081558.5 & 1 \\ -8539659.5 & 1 \end{bmatrix}$$

According to the formula (6), we have

$$\hat{\alpha} = [\alpha, \delta]^{T} = (\beta^{T}\beta)^{-1}\beta^{T}Z$$

Substituting relevant data into the above formula, we have

$$\hat{a} = \begin{bmatrix} \alpha \\ \delta \end{bmatrix} = \begin{bmatrix} 0.009 \\ 7.933 \end{bmatrix}$$

Then get the predicted value

$$\hat{w}^{(0)}(k+1) = \left[w^{(0)}(1) - \frac{\delta}{\alpha}\right] e^{-\alpha k} + \frac{\delta}{\alpha}, k = 0,1,2,\dots,n-1$$

$\hat{w}^{(0)}(1)$=-4286.92$e^{-(0.0.012\times 0)}$+4320.917
=34
$\hat{w}^{(0)}(2)$=-4286.92$e^{-(0.0.012\times 1)}$+4320.917
=85.13557

Evaluate of GM (1,1) forecasting model:

Table below show the precision of GM (1,1) forecasting model:

Table (7): Accuracy of the model

Test	GM (1,1) accuracy
Residual Test	0.206
c	0.176
MAPE	4.460
P=100-MAPE	95.940

Form table (1) it is clear that the GM (1,1) depending the above tests has a good model with acceptable accuracy. The specific results are shown in Table 8.

Table (8): Incidence of tuberculosis of Iraq from 1990 to 2023

Years	Fertility Rate $(\boldsymbol{w^{(0)}})$	$\hat{w}^{(0)}(k+1)$	$\hat{w}^{(k)}$
1990	34	34	1.00
1991	36	85.14	0.33
1992	38	135.66	0.38
1993	39	185.58	0.41
1994	40	234.91	0.45
1995	52	283.65	0.70
1996	59	331.81	0.41
1997	51	379.39	0.69
1998	51	426.41	0.66
1999	50	472.86	0.68
2000	50	518.76	0.65
2001	50	564.12	0.62
2002	49	608.93	0.64
2003	49	653.21	0.62
2004	49	696.96	0.59
2005	48	740.18	0.61
2006	48	782.89	0.59
2007	47	825.10	0.61
2008	47	866.80	0.59
2009	46	908.00	0.61
2010	45	948.71	0.64
2011	44	988.93	0.67
2012	43	1028.68	0.70
2013	41	1067.95	0.81
2014	39	1106.75	0.97
2015	37	1145.09	0.85
2016	35	1182.97	0.72
2017	33	1220.40	0.63
2018	31	1257.39	0.56
2019	28	1293.93	0.47
2020	26	1330.04	0.43
2021	24	1365.71	0.39
2022	23	1400.96	0.38
2023	42	1435.79	0.51

Grey Forecasting for fertility rate in Iraq during (1950-2023)

Table (9): Grey forecasting values for Incidence of tuberculosis of Iraq (2024-2027)

Years	Forecasting Incidence of tuberculosis of Iraq
2024	34
2025	34
2026	33
2027	33

Application of Grey-Markov Model for Incidence of tuberculosis of Iraq during the period (1990-2023)

State Partition: due to the small sample size in this study, the error can be divided into three states respectively, using E_1 , E_2 , E_3 as shown in the table below:

Table (10): State division for the error state

States	Interval
E_1	-0.18-0.06
E_2	0.06-0.29
E_3	0.29-0.53

From table (10), we obtained the error states for table 11

Table (11): Error states for the simulated value s Incidence of tuberculosis in Iraq during 1990- 2023

Years	Fertility Rate	$\hat{w}^{(0)}(k+1)$	w^k	Residual	Relative error	Error States
1990	34	34	1.00	0	0.00	1
1991	36	85.14	0.33	-15	0.42	3
1992	38	135.66	0.38	-13	0.33	3
1993	39	185.58	0.41	-11	0.28	2
1994	40	234.91	0.45	-9	0.23	2
1995	52	283.65	0.70	3	-0.06	1
1996	59	331.81	0.41	11	-0.18	1
1997	51	379.39	0.69	3	-0.07	1
1998	51	426.41	0.66	4	-0.08	1
1999	50	472.86	0.68	4	-0.07	1
2000	50	518.76	0.65	4	-0.08	1
2001	50	564.12	0.62	5	-0.09	1
2002	49	608.93	0.64	4	-0.09	1
2003	49	653.21	0.62	5	-0.10	1
2004	49	696.96	0.59	5	-0.11	1
2005	48	740.18	0.61	5	-0.10	1
2006	48	782.89	0.59	5	-0.11	1
2007	47	825.10	0.61	5	-0.10	1
2008	47	866.80	0.59	5	-0.11	1
2009	46	908.00	0.61	5	-0.10	1
2010	45	948.71	0.64	4	-0.10	1
2011	44	988.93	0.67	4	-0.09	1
2012	43	1028.68	0.70	3	-0.08	1
2013	41	1067.95	0.81	2	-0.04	1
2014	39	1106.75	0.97	0	-0.01	1
2015	37	1145.09	0.85	-1	0.04	1
2016	35	1182.97	0.72	-3	0.08	2
2017	33	1220.40	0.63	-4	0.13	2
2018	31	1257.39	0.56	-6	0.19	2
2019	28	1293.93	0.47	-9	0.31	3
2020	26	1330.04	0.43	-10	0.39	3
2021	24	1365.71	0.39	-12	0.49	3
2022	23	1400.96	0.38	-12	0.53	3
2023	42	1435.79	0.51	7	-0.17	1

From Table 11, We construct the transition Frequency Matrix (T.F.M)

	20	1	1	22
F=	1	3	1	5
	1	1	4	6
	22	5	6	33

Test for the existence of a Markov chain

To determine if the three-state transition chain adheres to the characteristics of a weighted Markov chain, we employ a Chi-square test as defined in equation (7):

Table (12): Chi-Square testing output of Incidence of tuberculosis in Iraq

	j = 1	j = 2	j = 3
i = 1	$\frac{\left(20-\frac{22\times22}{33}\right)^2}{\left(\frac{22\times22}{33}\right)}$ =1.939394	1.633333	2.25
i = 2	1.633333	6.637576	0.009091
i = 3	2.25	0.009091	7.757576
$\sum_{i=1}^{m}$	5.822727	8.28	10.01667
$\sum_{i=1}^{m}\sum_{i=1}^{n}$	24.11939		

The degrees of freedom are computed as (3-1)(3-1) = 4 when "m" equals 3. Using the Chi-square table with four degrees of freedom and a significance threshold of 5%, we find that$\chi^2_{(0.05,4)}$= 9.488. The Chi-square table value is less than the observed value of the sample statistic, chi-square, which is 24.11939. As a result, the hypothesis that states are independent is rejected. This demonstrates that the state transition chain for Iraq's TB incidence is a Markov chain. From (T.F.M.), we create the Transition Probability Matrix (T.P.M.).

	0.91	0.05	0.05
$P^{(1)}$ =	0.2	0.6	0.2
	0.17	0.17	0.67

The two steps, three steps, four steps up to the five steps the transition probability matrix is calculated as follows: -

	0.84	0.08	0.08
$P^{(2)}$ =	0.34	0.40	0.26
	0.30	0.22	0.49

	0.80	0.10	0.11
$P^{(3)}$ =	0.43	0.30	0.27
	0.39	0.23	0.38

To achieve this, we need to first find the error states of each of the year that is, from 1990 to 2023. From table 11, we observed that 2023 is in error state E_1 , this implies that, the initial state vector for the Grey-Markov prediction is

V_0	1	0	0

To obtain the error state for year 2024:

				0.91	0.05	0.05			
V_1=	1	0	0	0.2	0.6	0.2	0.91	0.05	0.05
				0.17	0.17	0.67			

It is observed from equation above that states are not equal probability of error state, however, we choose error state1(E1) for the 2024
To obtain the error state for year 2025

				0.84	0.08	0.08			
V_2=	0.91	0.05	0.05	0.34	0.4	0.26	0.7964	0.1038	0.1103
				0.3	0.22	0.49			

As it can be observed, error state 1 has the highest probability, therefore 2026 has error state1 (E1)

				0.8	0.1	0.11	0.8			
V_3=	0.7964	0.1038	0.1103	0.43	0.3	0.27	0.43	0.724771	0.136149	0.157544
				0.39	0.23	0.38	0.39			

Using equation $\hat{y}(k+1) = [1 + 1/2(R_{-i} + R_{+i})]\widehat{X_0}(k+1)$ and error states obtained for the respective years, we have the predictions from 2024 to 2028 respectively

$\hat{y}(2024) = [1 + 1/2(0.06 - 0.18)] \times 34 = 32$

$\hat{y}(2025) = [1 + 1/2(0.06 - 0.18)] \times 34 = 32$

$\hat{y}(2026) = [1 + 1/2(0.06 - 0.18)] \times 34 = 32$

$\hat{y}(2027) = [1 + 1/2(0.06 - 0.18)] \times 33 = 31$

6. Conclusions and Recommendations

A. Conclusion

In this study, the Bootstrap Grey Model GM (1,1) parameters were estimated and then used in a Hybrid Grey Markov Chain Model to predict the incidence of tuberculosis in Iraq. Among the important conclusions are:

(1) By resampling datasets to increase accuracy and robustness and lessen the influence of outliers and data variability, the Bootstrap Grey Model GM (1,1) improved upon the standard GM (1,1) and produced more reliable parameter estimations.

(2) Using a Markov Chain in conjunction with the Grey Model, the Hybrid model combined the Markov Chain's state-transition forecasting with the Grey Model's trend prediction capabilities. A more complete prediction framework that could handle both linear and non-linear data patterns was provided by this integration.

(3) The Hybrid Grey Markov Chain Model outperformed the GM (1,1) model by offering more precise and trustworthy forecasts for the occurrence of tuberculosis. The time series data's dynamic behavior was successfully recorded.

(4) Health data that can exhibit irregular patterns due to a variety of contributing factors, including seasonal variations and public health measures, such as the prevalence of tuberculosis, benefit greatly from this modeling method.

B. Recommendations

Based on the study's findings, several recommendations are proposed:

(1) For epidemiological forecasting, researchers and health authorities should use hybrid models such as the Hybrid Grey Markov Chain Model. These models help with improved planning and resource allocation by providing more precise predictions.

(2) New data should be added to the model on a regular basis to preserve accuracy. To reflect current developments, the Bootstrap Grey Model parameters and Markov Chain transition probabilities must be periodically recalibrated.

(3) To increase forecast accuracy, future models would benefit from include variables that affect the incidence of tuberculosis, such as socioeconomic factors, climate data, and public health initiatives.

(4) To improve proficiency in sophisticated modeling techniques and guarantee the efficient implementation and comprehension of these models, training programs for epidemiologists and health data analysts should be created.

(5) By using insights from these models, policymakers can prioritize high-risk locations highlighted by model projections for resource allocation and preventive actions, thereby implementing targeted interventions.

(6) It is recommended that more study be done to investigate combining Markov Chains and Grey Models with additional statistical and machine learning methods. Comparative research may be able to identify the best hybrid strategies for various kinds of epidemiological data.

References

1. Chen, C., & Hsu, C. (2006). Applying Bootstrap Method to Grey Forecasting Model. Journal of Grey System, 19(2), 87-92.
2. Deng, J. L. (1982). Control Problems of Grey Systems. Systems & Control Letters, 1(5), 288-294.
3. Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall/CRC.
4. Hamad, A. P. D. A. S., Faqe, A. P. D. M. M., & Mohamad, A. L. S. H. Forecasting Life-Expectancy in Iraq During the Period (2022-2035) Using Fuzzy Markov Chain.
5. Hamdin, A. O., & Hussein, M. M. (2023). An application of Wavelet Markov Chains Model to Study Earthquake Occurrence. Tikrit Journal of Administrative and Economic Sciences, 19(61), 2.
6. Hussein, M. M. F. (2023). Predicting Number of People Living with Chronic HCV Using Gray-Weighted Markov Chain Model (GW-MCM). University of Kirkuk Journal For Administrative and Economic Science, 13(4).
7. Hussein, M. M. F., Saeed, A. A., & Mohamad, S. H. (2023). Comparison Markov Chain and Neural Network Models for forecasting Population growth data in Iraq. University of Kirkuk Journal For Administrative and Economic Science, 13(4).
8. Li, S., & Wei, Y. (2016). A Hybrid Grey-Markov Forecasting Model Based on Bootstrap Resampling for Urban Traffic Flow. Journal of Traffic and Transportation Engineering, 3(3), 204-212.
9. Liu, S., & Lin, Y. (2006). Grey Systems: Theory and Applications. Springer.
10. Rashid, D. A., & Faqe, M. (2022). Constructing a Multilevel Modeling to High-Resolution CT (HRCT) Lung in Patients with COVID-19 Infection. Iraqi Journal of Statistical Sciences, 19(2), 78-90.
11. Wang, Y., Liu, S., & Dang, Y. (2018). A Hybrid Grey-Markov Chain Model for Energy Consumption Forecasting. Energy, 160, 593-601.
12. Yu, Y., Wang, H., & Lai, K. K. (2014). Hybrid Grey-Markov Prediction System in Financial Time Series. Expert Systems with Applications, 41(5), 1561-1568.
13. Zhang, H., Zhou, Y., & Liu, Y. (2019). Bootstrap-based Grey Forecasting Model and Its Application. Grey Systems: Theory and Application, 9(1), 12-25.
14. Zhou, M., Wang, X., & Wang, Y. (2020). Hybrid Grey-Markov Model for Air Quality Prediction. Environmental Monitoring and Assessment, 192(3), 1-17.