DOI: 10.24996/ijs.2024.65.5.28

ISSN: 0067-2904

Estimate the Parameters of the Weighted Exponential Regression Model for Panel Data

Hala Kadhum Obead [1], Saifaldin Hashim Kamar [2*], Basim Shlaibah Msallam [3]

[1]*Department of Banking and Finance, College of Management and Economics, Al-Iraqia University, Baghdad, Iraq*

[2, 3] *Department of Accounting, College of Management and Economics, Al-Iraqia University, Baghdad, Iraq*

Received: 14/3/2023 Accepted: 18/5/2023 Published: 30/5/2024

Abstract

The parameters of the weighted exponential regression model for panel data are estimated using the maximum likelihood method which represents the aim of this paper. Weekly infection and recovery ratios of COVID-19 data are predicted, where the model is converted from its nonlinear into a linear state using the Taylor series. The novelty of this paper lies in dealing with nonlinear panel data. Furthermore, the panel data of the model are tested to determine whether the data follows fixed or random effects by the Hausman test, as well as the exclusion of the pooled effects because the model does not include the intercept term. The simulation is depended on the generated data to compare the fixed and random effects models for different sample sizes (5, 10, 20, 30). COVID-19 data is used for three Iraqi governorates to represent the panel data model. Three months ,May, June, and July of 2022 are taken to represent the research sample and then predict the ratios of infection and recovery for the next three months. Depending on the Bayesian Information Criterion (BIC) and Akaike's Information Criterion (AIC), the random effects of the weighted exponential regression give better results than the fixed effects. Depending on this, we predict the weekly infection ratios of COVID-19 in Iraq that will decrease during the next ten weeks.

Keywords: Weighted exponential regression, Panel data, Taylor series, Growth curve, COVID-19.

تقدير معلمات نموذج الانحدار الأسي الموزون لبيانات اللوحة

حلا كاظم عبيد[1]، سيف الدين هاشم قمر[2*]، باسم شليبه مسلم[3]

[1] قسم العلوم المالية والمصرفية، كلية الإدارة والاقتصاد، الجامعة العراقية، بغداد، العراق.

[2، 3] قسم المحاسبة، كلية الإدارة والاقتصاد، الجامعة العراقية، بغداد، العراق.

الخلاصة

تم تقدير معلمات الانحدار الأسي الموزون لبيانات اللوحة باستخدام طريقة الإمكان الأعظم، والتي يمثل هدف البحث. وقد تم التنبؤ بنسب الإصابة والتعافي الأسبوعية اعتماداً على بيانات COVID-19، إذ تمت عملية تحويل النموذج من الحالة غير الخطية إلى الحالة الخطية باستخدام سلسلة Taylor، وتمثل أصالة هذا البحث في التعامل مع بيانات اللوحة غير الخطية. علاوة على ذلك، اختُبرت بيانات اللوحة الخاصة بالنموذج

* Email: dr.saifhkamar@gmail.com

لتحديد ما إذا كانت البيانات تتبع تأثيرات ثابتة أو عشوائية بواسطة اختبار Hausman، وكذلك استُبعد التأثيرات المجمعة لأن النموذج لا يتضمن الحد الثابت. وبالاعتماد على المحاكاة ولِّدت بيانات لغرض المقارنة بين نماذج التأثيرات الثابتة والعشوائية لأحجام عينات مختلفة (5، 10، 20، 30). وتم استخدام بيانات COVID-19 لثلاث محافظات عراقية لتمثيل نموذج بيانات اللوحة. إذ أُخذت ثلاثة أشهر (مايو ويونيو ويوليو) من عام 2022 لتمثيل عينة البحث ثم التنبؤ بنسب الإصابة والتعافي للأشهر الثلاثة القادمة. اعتمادًا على معيار معلومات بيز (BIC) ومعيار معلومات أكايكي (AIC)، أعطت التأثيرات العشوائية للانحدار الأسي الموزون نتائج أفضل من التأثيرات الثابتة. بناءً على ذلك، ظهرت معدلات الإصابة الأسبوعية بـ COVID-19 في العراق أنها ستنخفض خلال الأسابيع العشرة القادمة.

1. Introduction

The most important property that distinguishes the panel data from the cross-section is to give flexibility in modelling the differences in behaviour across individuals. The panel data has two dimensions, namely a time series and a cross-section. It is frequently used in economics and finance to analyze several questions that cannot be analyzed and explained using a time series or cross-section. For example, the importance of the exponential regression model in analyzing data from heterogeneous groups has been demonstrated. It proposed a simplified method using binary data to estimate the regression coefficients under a certain design of distributions, such as the elliptic symmetric distributions for explanatory variables, where the estimates are good, consistent, and normal, rivaling the maximum likelihood method when the sample size is large[1]. On the other hand, it has been proposed to estimate β for the exponential power distribution(EPD) based on the ordinary standard Q-Q scheme which is a direct method that does not require computational efforts exploiting the relationship between β and kurtosis[2]. Besides, the optimal D designs for linear and nonlinear models have been studied using weighted exponential functions and generalized exponential functions with symmetry variance. Therefore, these two functions can be applied as a growth curve. The results are obtained by employing the Kiefer-Wolowitz theorem[3].
Similarly, many of the structural properties of the bivariate and multivariate weighted exponential distribution have been studied, such as the second-order total positivity, marginal moments, reliability parameters, and estimated model parameters. The result of the simulation study is good and encouraging[4]. The performance estimator is also discovered of the least squares dummy variable (LSDV) for panel data models of small samples. It is compared with the performance of the intermediate variable of the generalized moment estimators (IV-GMM). Here, the authors used the properties of the Root Mean Squares Error (RMSE) of the model, the Root Mean Squares Error of the dependent variable (RMSEY) and Akaike's Information Criterion out of the confirmation of the usefulness of the estimator (LSDV) in determining the parameters of the dynamic panel model[5]. Two new twexp and twgravity methods for the weighted exponential regression models with constant effect are introduced, where they produce reliable standard errors. Note that this method is characterized by a fast calculation[6]. It reaches to study some statistical properties such as incomplete moment and moment, the mean of the residuals, and the survival function of a new distribution which is the log- weighted exponential (log-WE) distribution of the defined distribution during a specified period[7]. Here, unknown parameters of the proposed model are estimated by the method of the least squares and the weighted least squares and evaluated by studying Monte Carlo simulation experiments.

This research aims to estimate the parameters of the weighted exponential regression model for the panel data using the maximum likelihood method. The goal of this research is done according to the following. In the first section, some previous studies related to the research aim are reviewed. In the second section, the panel data concept and its four models are presented with the presentation of the weighted exponential regression model according to its original and minimum model and how to deal with it depending on the panel data models. Two models of panel data are discussed with a mention of the reasons.The third section relies on the study of simulation to achieve the aim of the research. Similarly, a practical application based on data on infection and recovery ratios with COVID-19 in Iraq and predicting those ratios in the coming months is presented in the fourth section. Finally, the fifth section concludes what has been reached.

2. Materials and Methods

2.1 Methods

The panel data has received much attention in the current decade, especially in the medical and economic fields, which simultaneously contain cross-sections and time series. It considers the impact of the change in time and cross-section[8]. Panel models come in four main forms, the pooled regression model, the fixed effects model, the random effects model, and the mixed effects model. In the case of the weighted exponential regression model, the fixed or random effects were tested using the Hausman test [9] which excludes the pooled effects because the model does not contain the fixed term. Thus, the hypothesis is as follows:

H_0: The random effects model rather than the fixed effects model is appropriate.

H_1: The random effects model is not appropriate.

If the p-value is less than 0.05, then we reject the null hypothesis; the model of constant effects is appropriate.

The parameters of the panel data model are estimated by the maximum likelihood method. Meanwhile, the Bayesian Information Criterion (BIC) and Akaike's Information Criterion (AIC) are used to verify the best model[10,11].

2.2 Weighted exponential regression model

Suppose the probability density function (pdf) of the weighted exponential regression model for the random variable *x* is given as follows[12]:

$$Y_i = \frac{\alpha+1}{\alpha}\lambda e^{-\lambda x_i}\left(1 - e^{-\alpha\lambda x_i}\right) + \varepsilon_i. \quad (1)$$

Hence, α $(\alpha > 0)$ is the shape parameter and λ $(\lambda > 0)$ is the scale parameter.

The weighted exponential model has been reformulated according to the following formula which is used in this work[3]:

$$Y_i = e^{-\alpha x_i}\left(1 - e^{-\lambda x_i}\right) + \varepsilon_i. \quad (2)$$

The substituting Eq. (2) yields:

$Y(X) = \emptyset(X, \theta) + \varepsilon.$

where $\theta = (\alpha,\ \lambda)$.

Using the Taylor series to $\emptyset(X, \theta)$ at the first derivative, we can write it as[13, 14]:

$$E(Y_i) = \emptyset(x_i, \theta) = \emptyset(x_i, \theta_0) + \sum_{j=1}^{2}\left(\frac{\partial\ \emptyset(x_i, \theta)}{\partial\theta_j}\bigg|\theta_j = \theta_{j0}\right)\left(\theta_j - \theta_{j0}\right),$$

where θ_{j0} are the initial values of the parameters.

By moving $\emptyset(X_i, \theta_0)$ to the left side of the equation, we get:

$$Y_i - \emptyset(X_i, \theta_0) = \sum_{j=1}^{2} X_{ij}^{*}\,\beta_{j0} + \varepsilon_i, i = 1,2, \ldots., n. \quad (3)$$

where $X_{ij}^{*} = \frac{\partial\ \emptyset(x_i,\theta)}{\partial\theta_j}|\theta_j = \theta_{j0}$ and $\beta_{j0} = \theta_j - \theta_{j0}$.

Thus, the model for the weighted exponential regression of panel models in terms of matrices is as follows:

$$Y_{it} = X'_{it}\beta + \varepsilon_{it}, \;\; t = 1, \dots, T, \;\; i = 1, \dots\dots, n. \tag{4}$$

where Y_{it} is the vector of the response variable with rank (n×1); X_{it} is the matrix of the explanatory variables with rank (2×n); β is the vector of parameters with rank (2×1); ε_{it} is the error term of rank (n×1) and T is the number of the time period.
Model (4) is free of an intercept. Therefore, we will not deal with pooled effects. Here, we will consider fixed and random effects to choose the appropriate estimate for the weighted exponential regression model through the Hausman test[9]. Although the regression models are written in linear form, the matrix of the explanatory variables can only be accessed by iterative methods. According to this, the model parameters can be estimated using the nonlinear maximum likelihood method.

2.3 Fixed effects model for the panel data

In general, any model of linear regression panel data has intercept terms and considers the change in the intercept terms for any section from one model to another for cross-section views in the sample. We can assume that the parameters change constantly. On this basis, they are called models of fixed effects. They represent both the individual and temporal dimension of the data panel model, in which the within estimator of the parameters is computed by one of the estimation methods, namely the maximum likelihood method:

$$\hat{\beta}_{FE} = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{Y}. \tag{5}$$

Here, $\hat{Y} = Y - \bar{Y}$ and $\hat{X} = X - \bar{X}$ are the transformed variables in deviations from the group mean. It is called a within estimator because it takes into account the variations in each group, then:

$Var(\varepsilon_{it}) = \sigma_\varepsilon^2, E(\varepsilon_{it}) = 0$, and $Var(\hat{\beta}_{FE}) = \sigma_{FE}^2(\hat{X}'\hat{X})^{-1}$, $\sigma_{FE}^2 = \frac{\varepsilon'\varepsilon}{nT-k}$,
where k is the number of explanatory variables.
However, we note that model (4) does not contain intercept terms. Hence, we neglect the model's fixed effects for intercept terms and try to the fixed effects with slopes without randomness. We only estimate slopes by the estimation method.

2.4 Random effects model for the panel data

This method randomly changes the intercept and slope terms since model (4) does not contain intercept terms. Subsequently, there is a random effect on the slope, and we obtain random coefficients regression model for panel data. Using the maximum likelihood method which is equivalent to the general least squares estimation. The final result for the estimated slope of the model (4) is given as follows:

$$\hat{\beta}_{RE} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}\ , \tag{6}$$

where $\tilde{X} = X - w\bar{X}$ and $\tilde{Y} = Y - w\bar{Y}$.

Accordingly to this, it has been suggested that [15]: $w = 1 - \sqrt{\frac{\sigma_{FE}^2}{\sigma_{FE}^2 + T\sigma_\mu^2}}$, and it has also been suggested that[16]: $\sigma_\mu^2 = \sigma_{POOLED}^2 - \sigma_{FE}^2$.

Consequently, $Var(\hat{\beta}_{RE}) = \sigma_{RE}^2(\tilde{X}'\tilde{X})^{-1}$, where $\sigma_{RE}^2 = \frac{\varepsilon'\varepsilon}{(N-k)}$, and σ_{RE}^2 has been suggested using the within regression residuals[17].

3. Simulation Study

It is certain that the weighted exponential regression model does not contain the intercept term after transforming it to the linear form, and it was mentioned in Section 2.1. Thus, the model is adopted with only random parameters (slopes), and it is also verified that the model is random by conducting the Hausman test.

Five simulation experiments are done to compare the fixed and random effects models described in Section 2. In these various experiments and depending on Eq. (2), we simulate the weekly recovery ratios of COVID-19 *(Y)* in three regions and we assume that (0.05, 0.03) are the initial values for the two parameters (α, λ) of the weighted exponential regression model with (5, 10, 20, 30) as a sample size of each region. The random error generation with zero mean and 0.5 variance is adopted. Here, the simulated model is replicated 500 times. In all experiments. Also, the estimation accuracy of the models is computed dependent on AIC and BIC.

Table 1: The results of the simulation

N	Criterion	Fixed effects	Random effects	Best
5	*AIC*	22.1595	20.5857	*Random effect*
	BIC	21.3784	19.8046	
10	*AIC*	121.7172	120.4434	*Random effect*
	BIC	122.3224	121.0486	
20	*AIC*	306.3382	279.3548	*Random effect*
	BIC	308.3297	281.3462	
30	*AIC*	382.2151	331.9353	*Random effect*
	BIC	385.0175	334.7377	

From Table 1, the random effects have the lowest AIC and BIC, so the random effects are better, which confirms what was stated in Section 2 of this research. We can also note that the best result is when $n = 5$, this proves the sensitivity of the weighted exponential regression model in the case of decreasing the sample size. This is the case for most growth models.

4. Real Data and Implementation

This section presents the data used by the research, which are represented in Table 2, and employed in estimating the weighted exponential regression model for panel data according to the targeted governorates Baghdad, Mosul, and Basra. We study the infection and recovery ratios of COVID-19, the ratio means the number of infection cases to the total number of PCR tests, during May, June, and July of 2022, knowing that the data is calculated for the weekly averages , the data is obtained from the Iraqi Ministry of Health[18].

Table 2: Infection and recovery ratios of COVID-19 for the governorates of Baghdad, Mosul, and Basra for May, June, and July of 2022

Governorate	Date	Weekly infection ratios	Weekly recovery ratios
Baghdad	08/05/2022	0.2926	0.3142
	15/05/2022	0.3433	0.2523
	22/05/2022	0.3969	0.3213
	29/05/2022	0.4027	0.3455
	05/06/2022	0.5046	0.4844
	12/06/2022	0.4742	0.5761
	19/06/2022	0.5567	0.5764
	26/06/2022	0.4960	0.6415
	03/07/2022	0.4127	0.5910
	10/07/2022	0.3485	0.5162
Mosul	08/05/2022	0.0072	0.0045
	15/05/2022	0.0072	0.0105
	22/05/2022	0.0109	0.0058
	29/05/2022	0.0054	0.0042
	05/06/2022	0.0065	0.0041
	12/06/2022	0.0014	0.0037
	19/06/2022	0.0021	0.0042
	26/06/2022	0.0080	0.0009
	03/07/2022	0.0099	0.0040
	10/07/2022	0.0177	0.0092

Basra	08/05/2022	0.0731	0.0664
	15/05/2022	0.0706	0.0686
	22/05/2022	0.0589	0.0502
	29/05/2022	0.0322	0.0323
	05/06/2022	0.0463	0.0380
	12/06/2022	0.0344	0.0284
	19/06/2022	0.0423	0.0436
	26/06/2022	0.0498	0.0215
	03/07/2022	0.0513	0.0576
	10/07/2022	0.0979	0.0572

Example 1. The weekly infection ratios of COVID-19
We assume the initial values of the parameter are (0.08, 0.02) for (b1, b2), respectively.

Table 3: Estimation of the parameters of the weighted exponential model depends on the weekly infection ratios of COVID-19

Model	α	λ	TD
Fixed effects	0.1989	0.0033	2.3203
Random effects	0.2401	0.0255	2.2779

We can notice, the random effects model gives the best TD (Total Deviation), where:

$$TD = \left|\frac{\hat{\alpha} - \alpha}{\alpha}\right| + \left|\frac{\hat{\lambda} - \lambda}{\lambda}\right|$$

And selects the best method using AIC and BIC, as shown in Table 4.

Table 4: AIC and BIC of fixed and random effects models for the weekly infection ratios of COVID-19

Model	AIC	BIC
Fixed effects	153.8128	154.4180
Random effects	127.7687	128.3739

Depending on the output of Table 4, we can select the random effects model because it gives the minimum AIC and BIC.

Table 5: Prediction of the weekly infection ratios of COVID-19

Week	Fixed effects	Random effects
1	0.00021	0.00032
2	0.00017	0.00026
3	0.00015	0.00021
4	0.00012	0.00017
5	0.00010	0.00013
6	0.00009	0.00011
7	0.00007	0.00008
8	0.00006	0.00007
9	0.00005	0.00005
10	0.00004	0.00004
Mean	0.00010	0.00010

According to Table 5, we can predict the average weekly infection ratios of COVID-19 in Iraq for the next ten weeks, with the best model being 0%.

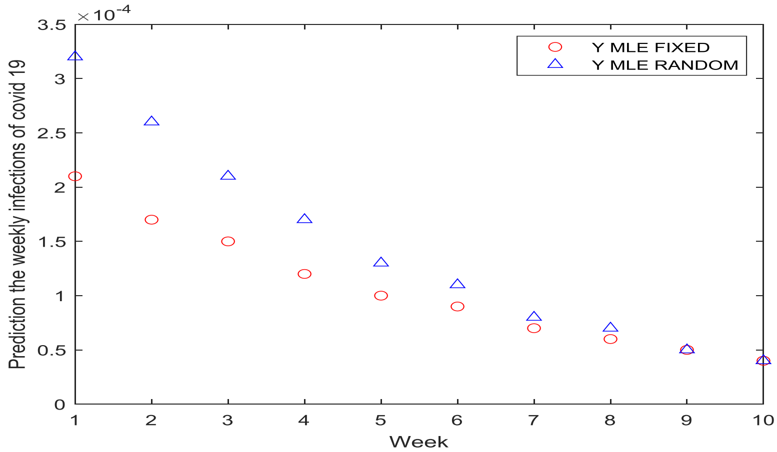

Figure 1: Prediction of the weekly infection ratios of COVID-19

It is clear from Figure 1 that the weekly infection ratios will decrease significantly during the next ten weeks to reach zero infection ratio per week.

Example 2. The weekly recovery ratios of COVID-19

We assume the initial values of the parameter are (0.06, 0.04) for (α, λ), respectively.

Table 6: Estimation of the parameters of the weighted exponential model depends on the weekly recovery ratios of COVID-19

Model	α	λ	TD
Fixed effects	0.0239	0.0126	1.2876
Random effects	0.0375	0.0379	0.4267

We can note that the random effects model gives the best TD and selects the best method using AIC and BIC, as shown in Table 7.

Table 7: AIC and BIC of fixed and random effects models for the weekly recovery ratios of COVID-19

Model	AIC	BIC
Fixed effects	69.8494	70.4546
Random effects	53.6384	54.2436

Depending on the output of Table 7, we can select the random effects model because it gives the minimum AIC and BIC.

Table 8: Prediction of the weekly recovery ratios of COVID-19

Week	Fixed effects	Random effects
1	0.15423	0.21631
2	0.15454	0.21181
3	0.15470	0.20723
4	0.15472	0.20258
5	0.15461	0.19788
6	0.15400	0.19316
7	0.15400	0.18800
8	0.15400	0.18400
9	0.15300	0.17900
10	0.15200	0.17400
Mean	0.15400	0.19540

According to Table 8, we can predict of average weekly recovery ratios of COVID-19 in Iraq, with the best model being 0.1954.

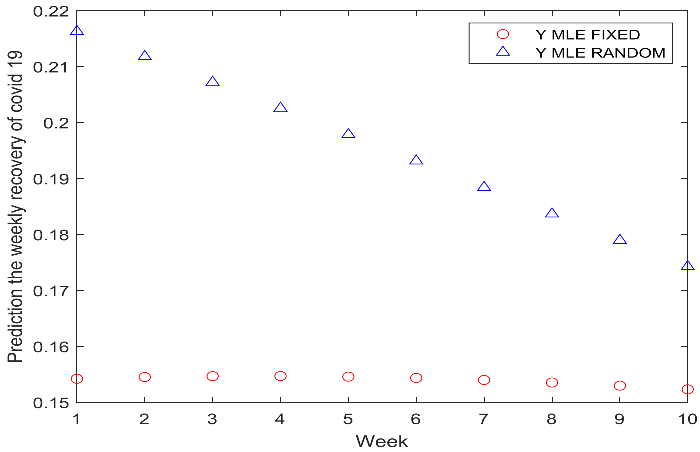

Figure 2: Prediction of the weekly recovery ratios of COVID-19

It is clear from Figure 2 that the weekly recovery ratios will decrease significantly during the next ten weeks to reach a 0.1740 infection ratio per week.

5. Conclusion

Depending on the simulation study, we conclude that the random effects method is better than the fixed effects method which is confirmed by the theoretical side of this research. In addition, it is also concluded that the weighted exponential regression model is more sensitive in the case of increasing the sample size, and it appears more efficient at small sample sizes. Relying on the real data for COVID-19 in Iraq, we can predict that the weekly infection ratios will decrease significantly during the next ten weeks to reach zero infection ratio per week with an average weekly infection ratio of 0.0001. Furthermore, it is also possible to predict the weekly recovery ratios that will decrease significantly during the next ten weeks to reach a 0.1740 infection ratio per week with an average weekly recovery ratio of 0.1954.

Acknowledgements

The authors gratefully acknowledge everyone who help them to complete this research.

References

[1] K. F. Cheng and J. W. Wu, "Estimation of exponential regression parameters using binary data," *Communications in Statistics-Theory and Methods*, vol. 21, no. 8, pp. 2203–2214, 1992.

[2] D. Coin, "A method to estimate power parameter in exponential power distribution via polynomial regression," *J Stat Comput Simul*, vol. 83, no. 11, pp. 1981–2001, 2013.

[3] T. Widiharih and S. Haryatmi, "D-optimal designs for weighted exponential and generalized exponential models," *Applied Mathematical Sciences*, vol. 7, no. 22, pp. 1067–1079, 2013.

[4] I. Ghosh and A. Alzaatreh, "On the bivariate and multivariate weighted generalized exponential distributions," *Hacettepe Journal of Mathematics and Statistics*, vol. 45, no. 5, pp. 1525–1540, 2016.

[5] J. U. Okeke and E. N. Okeke, "Least squares dummy variable in determination of dynamic panel model parameters," *European Journal of Engineering and Technology Research*, vol. 1, no. 6, pp. 77–81, 2016.

[6] K. Jochmans and V. Verardi, "Fitting exponential regression models with two-way fixed effects," *Stata J*, vol. 20, no. 2, pp. 468–480, 2020.

[7] E. Altun, "The log-weighted exponential regression model: alternative to the beta regression model," *Communications in Statistics-Theory and Methods*, vol. 50, no. 10, pp. 2306–2321, 2021.

[8] B. H. Baltagi, *The Oxford handbook of panel data*. Oxford Handbooks, 2015.
[9] J. A. Hausman, "Specification tests in econometrics," *Econometrica*, pp. 1251–1271, 1978.
[10] J. Brownlee, "Probabilistic model selection with AIC, BIC, and MDL," *Probability; Machine Learning Mastery: San Francisco, CA, USA*, 2019.
[11] A. A. Neath and J. E. Cavanaugh, "The Bayesian information criterion: background, derivation, and applications," *Wiley Interdiscip Rev Comput Stat*, vol. 4, no. 2, pp. 199–203, 2012.
[12] P. Nasiri, I. Makhdoom, and B. Yaghoubian, "Estimation parameters of the weighted exponential distribution," *Aust J Basic Appl Sci*, vol. 5, no. 9, pp. 2007–2014, 2011.
[13] S. H. Kamar, "Bayesian Procedure Modified by Fourier Series Residual to Fitting the Logistic Growth Model," *Pak. J. Statist*, vol. 35, no. 4, pp. 301–313, 2019.
[14] S. H. Kamar, B. S. Msallam, and H. S. Uraibi, "Ordinary Least Squares based on Fourier Series Residuals for fitting the Modified Exponential Growth Model," in *Journal of Physics: Conference Series*, 2021, vol. 1897, no. 1, p. 12001.
[15] I. C. Álvarez, J. Barbero, and J. L. Zofío, "A panel data toolbox for Matlab," *J Stat Softw*, vol. 76, pp. 1–27, 2017.
[16] W. H. Greene, *Econometric analysis*. Pearson Education India, 2003.
[17] P. Swamy and S. S. Arora, "The exact finite sample properties of the estimators of coefficients in the error components regression models," *Econometrica*, pp. 261–275, 1972.
[18] Iraqi Ministry of Health, "No Title," 2022.