Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:121
https://doi.org/10.28919/cmbn/8795
ISSN: 2052-2541

PROPOSED ROBUST ESTIMATORS FOR THE POISSON PANEL REGRESSION MODEL: APPLICATION TO COVID-19 DEATHS IN EUROPE

ELSAYED G. AHMED[1], MOHAMED R. ABONAZEL[2,*], MOHAMMED NAJI AL-GHAMDI[3], HANY M. ELSHAMY[4], IBRAHIM G. KHATTAB[5]

[1]Department of Statistics and Insurance, Faculty of Commerce, Arish University, North Sinai, Egypt
[2]Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt
[3]Department of General Studies, Technical college of telecom and information - Technical and Vocational Training Corporation (TVTC), Jeddah, Saudi Arabia
[4]Department of Economics, Faculty of Commerce, Tanta University, Tanta, Egypt
[5]Department of Statistics, Mathematics, and Insurance, Faculty of Business, Alexandria University, Alexandria, Egypt

Abstract. In regression panel data analysis, the maximum likelihood (ML) estimates of the Poisson model with fixed effects (FE) are affected by outliers. Thus, the ML estimation method will not be appropriate to solve the problem of outliers in the panel data. Therefore, we need robust estimation methods where the estimates of these methods are not much affected when the dataset contains outliers. This study aims to propose three robust estimation (M, S, and MM) methods that deal with panel datasets that contain outliers to enhance the accuracy of the results and provide good, stable, and more accurate predictions. For this purpose, these proposed robust methods were applied to coronavirus data for twelve high-income countries in Europe during the period from June 23, 2021, to January 21, 2022, to examine the performance and efficiency of these estimators in the presence of outliers. The results of COVID-19 indicated that the estimates of the classical ML estimation method are highly

*Corresponding author
E-mail address: mabonazel@cu.edu.eg
Received July 30, 2024

sensitive to outliers unlike proposed robust estimation methods, especially the MM robust estimation method, where the MM estimates are better than the other estimates.

Keywords: count panel data; outliers; negative binomial panel; fixed effects; maximum likelihood; M fixed effects Poisson; S fixed effects Poisson; MM fixed effects Poisson.

2020 AMS Subject Classification: 62J12, 62J20, 62F35.

1. INTRODUCTION

Over the last few years, panel data models have gained increasing importance in different fields, such as econometrics, as these models take the unobserved individual heterogeneity into account. Panel dataset refers to the two-dimensional dataset in which cross-section units are observed over some time. In this study, we will focus on the most popular model among researchers in panel data regression models which is the FE model. In the FE panel data model, there is an intercept term for each cross-sectional unit i. For more details about panel data regression models, see [1, 2, 3, 4]. A static linear FE panel data model can be represented as:

$$y_{it} = \eta_i + x_{it}^T \beta + u_{it}, \quad i = 1,2,\ldots,N; \quad t = 1,2,\ldots,T \tag{1}$$

where y_{it} denotes the dependent variable, $x_{it} \in \mathbb{R}^K$ denotes the independent variables which have the K-dimensional and $\beta \in \mathbb{R}^K$ is the vector of the regression parameters. The subscript i denotes the number of individuals, households, and countries or firms over a period from time t. While the η_i represents the unobservable individual effects (scalar constant) and u_{it} is the independent and identically distributed error term.

The linear FE panel data model in (1) assumes a normal distribution for the dependent variable. But sometimes, we find that the dependent variable in linear panel data regression does not follow the normal distribution, so model (1) is not appropriate for the count panel data, where the dependent variable takes positive integer values. Thus, linear estimation methods, like least squares are not suitable for count data, as these methods are specifically designed for continuous variables. So, count data analysis relies on distributions like the fixed effects Poisson (FEP) panel model, which provides a more suitable basis for handling such discrete data. These distributions take into account the nature of count data and offer more accurate and meaningful results for analyzing and interpreting such data, see [5, 6, 7].

2. CLASSICAL FEP PANEL DATA MODEL

The FEP model is one of the most widely used models in econometric models. When the response variable is a non-negative integer, this model can control for unobserved heterogeneity that could lead to biased estimates. Many researchers in econometrics are based on count panel data models because traditional ordinary least squares (OLS) regression techniques do not fit for count dataset, as the OLS assumes that the dependent variable is continuous and follows normal distribution. In this context, the FEP model is used as one of the most important count panel data models to deal with the data of the dependent variable that takes non-negative integer values.

The FEP regression model is used to analyze panel data, particularly when dealing with count data. This model extends the traditional Poisson regression with accounting for individual-specific effects, also known as FE. It is commonly applied in various fields such as economics, public health, and social sciences. This model enables researchers to gain deeper insights into the impact of explanatory variables on the occurrence of events or counts by accounting for individual-specific differences that may influence the outcomes.

Also, the FEP model allows for examination of the relationship between a count-based response variable, distributed according to the Poisson distribution, and different explanatory variables with controlling for individual-specific characteristics that remain constant over time, this model addresses unobservable individual heterogeneity, which is crucial when dealing with panel data where observations are repeated across different individuals or units over time, see [8, 9, 10]. The probability mass function of the FEP regression panel data model is:

$$f(y_{it}^p|x_{it},\eta_i,\beta) = \frac{\exp(-\mu_{it})}{y_{it}^p!} \times (\mu_{it})^{y_{it}^p}, \quad i = 1,2,\ldots,N; \quad t = 1,2,\ldots,T \tag{2}$$

where y_{it}^p is the dependent variable, which takes positive integers ($y_{it}^p \geq 0$) for units i at time t. The superscript p indicates that the variable y_{it} follows a Poisson distribution, where the mean and the variance of the dependent variable are equal in (2).

The fixed effects negative binomial (FENB) panel model is sometimes utilized as an alternative to the Poisson panel model to deal with the problem of overdispersion in real datasets. This is because it allows the mean of the response variable to differ from its variance, see [11, 12].

In this context, we introduce robust methods that deal with the nature of data that contain outliers using the FE model for Poisson by the application on a set of panel data for COVID-19 to understand the patterns and spread of the virus and provide valuable insights about the behavior and trend of the virus. This provides public health officials and policymakers with insights into the pandemic's dynamics, enabling them to devise effective strategies and policies for managing and mitigating the impact of the pandemic on community and public health in the future.

3. Robust Estimators of FEP Panel Model

Scientific research and statistical analyses sometimes deal with panel datasets that contain outliers, which can affect the accuracy of results and predictions. In this context, the importance of employing robust regression in data analysis is important. The robust regression is considered a vital tool in the field of statistical analysis, as it aims to provide stable and reliable analytical results in the presence of outliers in the data. This type of analysis relies on statistical procedures that provide estimates of relationships among variables without being significantly affected by outliers or values that fall outside the general pattern of the data. An important aspect of using robust regression is that it contributes to reducing the influence of outliers, thereby yielding more stable and accurate results for forecasts and analyses. Additionally, robust regression works to enhance the analytical models' capability to handle variability in the data, making them more effective in interpreting relationships between variables.

The traditional estimation methods of the Poisson model with FE are highly sensitive when outliers are present in the panel data. Hence, these methods are expected to produce non-robust estimates. Therefore, we propose robust estimates for the FEP panel data model. These robust estimates can combine robustness and efficiency to obtain resistant and robust results if the panel data contains outliers.

Several robust estimation methods have been introduced to achieve robustness and efficiency in various regression models, see e.g. [13, 14, 15, 16, 17, 18, 19, 20, 21]. However, there is no robust method to estimate the parameters in the FEP panel data model. In Section (3), we will propose robust estimation methods for the Poisson model with FE. Furthermore, a detailed description of the algorithms of these methods will be presented.

Based on the Poisson panel data model with FE in (2), we find that the joint probability function (JPF) for the i^{th} observation in a model (2) is:

$$f(y_{it}^{p}|\eta_i,\beta,x_{it}) = \prod_{t=1}^{T}\left(\frac{\exp(-\mu_{it})}{y_{it}^{p}!}\times(\mu_{it})^{y_{it}^{p}}\right) \tag{3}$$

where $\mu_{it} = \exp(\omega_i + x_{it}'\beta)$ and ω_i is logarithm η_i. Taking the logarithm of JPF in (3) for the i^{th} observation and summing over all cross-section units, the log-likelihood function of unit i is:

$$l(\eta_i,\beta) = -\eta_i T\bar{\lambda}_i + \omega_i T\bar{y}_i + T\bar{y}_i \ln\lambda_{it} - \sum_{t=1}^{T}\ln(y_{it}!) \tag{4}$$

where $T\bar{\lambda}_i = \sum_{t=1}^{T}\lambda_{it}$ and $T\bar{y}_i = \sum_{t=1}^{T} y_{it},$. From (4), we get the ML estimator for individual effects of cross-sections η_i by solving the log-likelihood function in (4) as follows:

$$\eta_i = \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T}\lambda_{it}} \tag{5}$$

We can get the first-order condition for β from (4) independently of η_i after substituting individual effects obtained in (5) and the first derivative, in addition to setting the equation equal to zero as:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}'\left[y_{it} - \left(\frac{\bar{y}_i}{\bar{\lambda}_i}\right)\lambda_{it}\right] = 0 \tag{6}$$

It can obtain coefficients of the FEP for β by solving (6). We can write the disturbances of the Poisson panel data model with FE in (2) as follows:

$$D_{it}(\eta_i,\beta) = y_{it} - \exp(\omega_i + x_{it}'\beta) \tag{7}$$

It can look at the value of the residuals to judge the suitability of the estimated regression model to the data. Residuals are most used as measures to detect outliers, where the points that are found far from the line and have large residual values are known as outliers [22].

The robust estimation methods are an important approach to dealing with the count panel data which contain outliers. In the FEP model, using robust estimates becomes necessary when detecting outliers to provide reliable and stable results in the case presence of outliers in the count panel data. The main objective of this paper is to propose robust FEP estimators that can resist the expected damaging impact of outliers in the count panel data. To achieve these objectives, we will discuss three robust estimators for the FEP model in the presence of outliers and compare their performance with the ML estimator (non-robust).

3.1. M Fixed Effects Poisson Estimation.

M fixed effects Poisson (MFEP) estimation is an extension of the ML estimation method and a robust estimation for the parameters of the FEP model. The MFEP estimator is based on minimizing the disturbance function $\rho(.)$, see, e.g., [22, 23, 24]. In this context, we can define the MFEP estimator of β depending on the function $\rho(.)$ by minimizing the disturbance function as follows:

$$\hat{\beta}_{MFEP} = \min \sum_{i=1}^{N} \sum_{t=1}^{T} \rho(u_{it}^{M}) \tag{8}$$

The disturbances in (7) are standardized by using a measure of dispersion $\hat{\sigma}$. Therefore, standardized disturbances u_{it}^{M} in (8) equal $\frac{D_{it}^{M}(\eta_i,\beta)}{\hat{\sigma}_{SQAD}}$. We can set an estimator for $\hat{\sigma}$ to be the second quartile absolute deviation (*SQAD*) as follows:

$$\hat{\sigma}_{SQAD} = F \times SQ\left(\left|D_{it}^{M}(\eta_i,\beta) - SQ\left(D_{it}^{M}(\eta_i,\beta)\right)\right|\right) \tag{9}$$

where SQ is the second quartile and the estimator $\hat{\sigma}_{SQAD}$ rescales $SQAD$ by the factor F which is 1.4826. For the ρ function in (8), we can rewrite the objective function of Tukey's bisquare by using panel data as follows:

$$\rho(u_{it}^{M}) = \begin{cases} \frac{(u_{it}^{M})^2}{2} - \frac{(u_{it}^{M})^4}{2c^2} + \frac{(u_{it}^{M})^6}{6c^4}, & |u_{it}^{M}| \leq c \\ \frac{c^2}{6}, & |u_{it}^{M}| > c \end{cases}$$

We can take an appropriate value of the tuning constant c to achieve a reasonably high level of efficiency. Usually, the smaller value for c gives more resistance in the presence of outliers, see, [25, 26, 27, 28]. With the first partial derivative for (8) and setting the partial derivatives to zero, we can obtain:

$$\min_{\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} \psi\left(\frac{D_{it}^{M}(\eta_i,\beta)}{\hat{\sigma}}\right)\left(\frac{\partial(D_{it}^{M}(\eta_i,\beta))}{\partial \beta_j}\right) = 0, j = 1,\ldots,K \tag{10}$$

$\psi(u_{it}^{M})$ is referred to as the influence function, which is $\rho'(u_{it}^{M})$. It can be written (10) using the weight function as follows:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(D_{it}^{M}(\eta_i,\beta)\right)' W_M(D_{it}^{M}(\eta_i,\beta))D_{it}^{M}(\eta_i,\beta)=0 \tag{11}$$

where $\left(D_{it}^{M}(\eta_i,\beta)\right)' = \left[\frac{\partial(D_{it}^{M}(\eta_i,\beta))}{\partial\beta_j}\right]$ and $W_M(D_{it}^{M}(\eta_i,\beta)) = \left[\frac{\psi(u_{it}^{M})}{u_{it}^{M}}\right]$ is a matrix $(NT\times NT)$ which the diagonal elements are weighted. Solution for (11) gives robust estimators of the FEP model. We will show a detailed description of M robust estimation for the FEP in Algorithm (3.1.1).

3.1.1. *Algorithm of MFEP Estimation.*

(1) Estimate parameters for the FEP model using the ML estimation method.
(2) Calculate initial regression parameters $\hat{\beta}_{MFEP}^{0}$ with the FEP model.
(3) Calculate disturbances value $D_{it}^{M}(\eta_i,\beta)$ of MFEP.
(4) Calculate second quartile absolute deviation $\hat{\sigma}_{SQAD}$.
(5) Calculate the value of standardized disturbances u_{it}^{M}.
(6) Calculate the weighted value as follows:

$$W_M(D_{it}^{M}) = \begin{cases} [1-\left(\frac{u_{it}^{M}}{c}\right)^2]^2 & \text{if } |u_{it}^{M}|\le c \\ 0 & \text{if } |u_{it}^{M}|> c \end{cases}$$

(7) Estimate coefficients of MFEP ($\hat{\beta}_{MFEP}$) by using the ML method with $W_M(D_{it}^{M})$.
(8) Repeat the steps from (3) to (6) to obtain a convergent value of the $\hat{\beta}_{MFEP}$.
(9) Examine the results and performance of the estimators by using some criteria.

3.2. S Fixed Effects Poisson Estimation. In this section, we propose S fixed effects Poisson (SFEP) estimation which is associated with MFEP estimation. We will modify the S estimation method which was suggested by [29]. The SFEP estimation method depends on the disturbances scale of the MFEP estimation method by minimizing the scale of dispersion of the disturbances. In this method, we will use the disturbance standard deviation to overcome the weaknesses of the second quartile, see [22]. It can define the SFEP as follows:

$$\hat{\beta}_{SFEP} = \min \hat{\sigma}_{SFEP}(D_{11}^{S}(\eta_i,\beta),\ldots,D_{NT}^{S}(\eta_i,\beta)) \tag{12}$$

According to [29] and [30], it can increase robustness by obtaining the smallest robust scale of the disturbances $\hat{\sigma}_{SFEP}$ and satisfying:

$$\min \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\rho \left(\frac{D_{it}^{S}(\eta_i, \beta)}{\hat{\sigma}_{SFEP}} \right) \right] \tag{13}$$

The solution of (13) is obtained by using differentiating of the coefficients as follows:

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \psi \left(\frac{D_{it}^{S}(\eta_i, \beta)}{\hat{\sigma}_{SFEP}} \right) \left(\frac{\partial D_{it}^{S}(\eta_i, \beta)}{\partial \beta_j} \right) = 0 \tag{14}$$

where ψ is a function that represents the derivative of ρ as follows:

$$\psi(u_{it}^{S}) = \rho'(u_{it}^{S}) = \begin{cases} u_{it}^{S} \left[1 - \left(\frac{u_{it}^{S}}{c} \right)^2 \right]^2 & \text{if } |u_{it}^{S}| \leq c \\ 0 & \text{if } |u_{it}^{S}| > c \end{cases}$$

$u_{it}^{S} = \left(\frac{D_{it}^{M}(\eta_i, \beta)}{\hat{\sigma}_{SFEP}} \right)$. The equation (15) can be solved by using iteratively reweighted ML method. In Algorithm (3.2.1), we will show stages of S robust estimation for the FEP model.

3.2.1. *Algorithm of SFEP Estimation.*

(1) Estimate parameters for the FEP model using the ML estimation method.

(2) Calculate initial regression parameters ($\hat{\beta}_{SFEP}^{0}$) with MFEP.

(3) Calculate the disturbances value of SFEP $D_{it}^{S}(\eta_i, \beta)$.

(4) Calculate value $\hat{\sigma}_{SFEP}$ as follows:

$$\hat{\sigma}_{SFEP} = \begin{cases} \hat{\sigma}_{SQAD} & \text{, iteration = 1;} \\ \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[W_M(D_{it}^{S})^2 \right]} & \text{, iteration} > 1. \end{cases}$$

(5) Calculate the value of standardized disturbances u_{it}^{S}.

(6) Calculate the weighted value as follows:

$$W_S(D_{it}^{S}) = \begin{cases} \begin{cases} \left[1 - \left(\frac{u_{it}^{S}}{c} \right)^2 \right]^2, & |u_{it}^{S}| \leq c \\ 0, & |u_{it}^{S}| > c \end{cases} & \text{, iteration} = 1; \\ \left(\frac{\rho(u_{it}^{S})}{(u_{it}^{S})^2} \right), & \text{iteration} > 1. \end{cases}$$

and

$$\rho(u_{it}^S) = \begin{cases} \frac{(u_{it}^S)^2}{2} - \frac{(u_{it}^S)^4}{2c^2} + \frac{(u_{it}^S)^6}{6c^4} & \text{if } |u_{it}^S| \leq c \\ \frac{c^2}{6} & \text{if } |u_{it}^S| > c \end{cases}$$

(7) Estimate coefficients of SFEP ($\hat{\beta}_{SFEP}$) by using $W_S(D_{it}^S)$.

(8) Repeat the steps from (3) to (6) to obtain a convergent value of the $\hat{\beta}_{SFEP}$.

(9) Examine the results and performance of the estimators by using some criteria.

3.3. MM Fixed Effects Poisson Estimation. The MM fixed effects Poisson (MMFEP) method is based on estimating the regression coefficients by using SFEP estimation which minimizes the scale of the disturbance from MFEP estimation. The MMFEP estimation method aims to get estimates with a higher breakdown value and more efficiency, see e.g. [31]. It can obtain MMFEP estimator by solving:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \psi\left(\frac{D_{it}^{MM}(\eta_i,\beta)}{\hat{\sigma}_{MMFEP}}\right)\left(\frac{\partial D_{it}^{MM}(\eta_i,\beta)}{\partial \beta_j}\right) = 0 \tag{15}$$

where $\psi(u_{it}^{MM}) = \rho'(u_{it}^{MM})$ and $\hat{\sigma}_{MMFEP}$ represents the standard deviation calculated from the disturbances of the SFEP estimation method. In Algorithm (3.3.1), we will provide a detailed description of MM robust estimation for the FEP model.

3.3.1. *Algorithm of MMFEP Estimation.*

(1) Estimate parameters for the FEP model using the ML estimation method.

(2) Calculate initial regression parameters ($\hat{\beta}_{MMFEP}^0$) with SFEP.

(3) Calculate the disturbances value of MMFEP $D_{it}^{MM}(\eta_i,\beta)$.

(4) Calculate value $\hat{\sigma}_{MMFEP}$.

(5) Calculate the value of standardized disturbances $u_{it}^{MM} = \left(\frac{D_{it}^{MM}(\eta_i,\beta)}{\hat{\sigma}_{MMFEP}}\right)$.

(6) Use objective function $\rho(u_{it}^{MM})$ and u_{it}^{MM} to calculate the $W_{MM}(D_{it}^{MM})$.

(7) Estimate coefficients of MMFEP ($\hat{\beta}_{MMFEP}$) by using $W_{MM}(D_{it}^{MM})$.

(8) Repeat the steps from (3) to (6) to obtain a convergent value of the $\hat{\beta}_{MMFEP}$.

(9) Examine the results and performance of the estimators by using some criteria.

The use of robust estimates of FEP is fundamental in scientific research and statistical analyses, as it contributes to improving the quality of results and recommendations. It reduces the influence of outliers in the analysis of relationships between variables, thereby contributing significantly to more reliable and robust findings.

4. Application on Cases of COVID-19 Infection

In 2019, the COVID-19 pandemic has captured extensive attention due to its widespread impact on global economies and the immense loss of lives it has caused. Because of the absence of antiviral drugs or vaccines, the incidence of new cases of coronavirus has surged dramatically, leading to innumerable deaths. Consequently, there has been a vital focus on developing diverse methodologies to analyze pandemic data, particularly for predicting future cases of the coronavirus. Naturally, many statistical studies have also been developed to study and predict the number of people infected with or die from Coronavirus [32] or other epidemic diseases [33].

[34] employ panel data to model and analyze the dynamics of COVID-19 infected cases. The research focuses on applying statistical modeling techniques to comprehend the patterns and spread of COVID-19 cases. The study aims to provide valuable insights into the behavior and trends of the COVID-19 pandemic by utilizing panel data models, contributing to a better understanding of its statistical aspects. In this paper, the panel data dataset for COVID-19 cases was collected from daily bulletins in four districts of India (Kanniyakumari, Tenkasi, Thoothukudi, and Tirunelveli) during the period from October 1, 2020, to October 31, 2020. Panel data models were utilized to analyze the trends, where the number of new cases was the dependent variable and time was the independent variable. The panel data regression models were found to be more appropriate than classical models.

[35] focused on analyzing the effects of COVID-19 data with panel data and some statistical methods to delve into the complex and evolving nature of the pandemic. The researchers utilized this approach to study the effects of various factors related to COVID-19, such as number of cases, number of tests, stringency index, population, average age, number of beds in the hospital, and gross domestic product on the number of deaths caused by the COVID-19 virus for 20 countries in the period from March 12, 2020, to May 29, 2020. The study aimed to understand the interplay and dynamics of COVID-19 and its implications on public health and management strategies. The study's findings provide valuable insights for policymakers and

public health officials into the dynamics of the pandemic and can inform strategies and policies to effectively manage and mitigate its impact on public health and society.

[36] applied the panel data models (pooled, random effects, and FE) to analyze the impact of government policies and coronavirus cases on people's mobility and activity participation in seventeen Asian countries (Afghanistan, Bangladesh, India, Iraq, Indonesia, Japan, Myanmar, Pakistan, Philippines, South Korea, Saudi Arabia, Singapore, Sri Lanka, Thailand, Turkey, United Arab Emirates, and Vietnam) during 300 days in the period from February 15, 2020, to December 10, 2020. The paper analyses the effect of three independent variables: the stringency index, reproduction number, and economic support index on mobility and activity participation, there are six types of activities: retail and recreation, workplaces, grocery and pharmacy, park, transit station, and residential. The research aims in general to shed light on the effectiveness of different policies implemented by governments and how they correlate with mobility trends and activity levels during the pandemic. This analysis is crucial in assessing the degree to which public health measures and government interventions impact the daily lives of individuals in terms of movement and participation in activities, especially during a prolonged period of 300 days. By evaluating these aspects across different Asian countries, the study aims to provide valuable insights and contribute to a better understanding of the relationship between policies, COVID-19 cases, mobility, and activity participation during the ongoing pandemic.

The COVID-19 pandemic that started in Wuhan, China, in late 2019 influenced all countries of the world and caused a global economic crisis whose effects will remain for years. This requires continuous monitoring and forecasting of COVID-19 prevalence to ensure effective control. In this regard, we evaluate the robust methods that were introduced in this paper by applying them to cases of COVID-19 infection and some related independent variables.

4.1. Data Description. In this application, we used the daily data for twelve high-income countries in Europe according to the World Bank classification during the period from June 23, 2021, to January 21, 2022, from the World Health Organization website. R, STATA, and E-views software were used to get the results in this application. Table 1 displays the definition of the variables used in this study.

TABLE 1. Definition of Study Variables

Variables	**Description**
Dependent	
Deaths Cases	Number of new deaths cases for COVID-19 (Count)
Independent	
Infected Cases	Number of new confirmed cases of COVID-19 per 10,000 people
Vaccinated Persons	The logarithm of new COVID-19 vaccinations administered to people
Tested Cases	The logarithm of the Number of new tests for COVID-19
Positive Rate	The share of COVID-19 tests that are positive (%)

Table 2 presents some of the descriptive statistics of the five variables in this study. The average number of new death cases for COVID-19 in all countries was 17.3271, with a maximum of 434 cases. While the average number of confirmed cases was 0.5312 per 10,000 people and the value maximum of infections was 22.8123. Regarding the average of the vaccinated persons, tested cases, and positive rate for these tests were 8.4129, 10.8411, and 0.0666 respectively.

TABLE 2. Descriptive Statistics of Study Variables ($N \times T = 2556$)

Variables	Mean	Std. Dev.	Median	Min.	Max.
Deaths Cases	17.3271	33.3264	5	0	434
Infected Cases	0.5312	1.6717	0.1457	0.0003	22.8123
Vaccinated Persons	8.4129	1.5330	8.3790	4.0775	12.5733
Tested Cases	10.8411	1.1720	10.6873	8.0956	14.6023
Positive Rate	0.0666	0.0834	0.0350	0.0002	0.5100

Table 3 presents some summary statistics for countries under study. The results reveal that country of Italy contains the largest number of cases infected with COVID-19, with the number of infections reaching approximately 434, while the largest number of COVID-19 infections in other countries ranges from the number of infections in the country of Cyprus, which was registered 11 cases, and the country of Czechia, which was registered 140 cases.

TABLE 3. Some Descriptive Statistics for Death Cases of every Country

Cross Section	Country Name	Mean	Median	Variance	Min.	Max.
1	Austria	18.892	10	443.767	0	80
2	Belgium	16.545	10	226.013	0	61
3	Cyprus	1.563	1	3.2	0	11
4	Czechia	31.408	4	1727.856	0	140
5	Denmark	4.784	3	42.038	0	60
6	Finland	5.202	3	32.624	0	26
7	Ireland	5.967	6	18.041	0	18
8	Italy	71.826	49	5577.201	2	434
9	Lithuania	15.779	17	134.503	0	44
10	Norway	3.347	0	89.03	0	59
11	Slovakia	23.826	9	831.22	0	113
12	Switzerland	8.784	6	73.821	0	35

4.2. Outliers Diagnostics. Detecting outliers is a crucial process in regression analysis and it is a fundamental step in various dataset applications. Box plot has been employed to detect outliers of each variable. Fig. 1 displays a box plot for five variables under study. These figures show that each variable in this study has outliers.

If the dataset contains outliers, then the classical estimates (non-robust) are not efficient, robust estimators should be used to obtain regression coefficients, see e.g. [37, 38, 17].

4.3. Unit Root Test for the used Variables. It is necessary to conduct unit root tests for variables under study before estimating and analyzing the panel data regression models to examine the stationary of the variables because estimating non-stationary data may lead to misleading results and conclusions. Also, the panel data are susceptible to instability and volatility due to the presence of individual variations and differences among cross-section units in the panel data [39].

The results of unit root tests are presented in Table 4 confirms that the null hypothesis for all variables has been rejected. This indicates that the five variables are stationary at < 0.0001

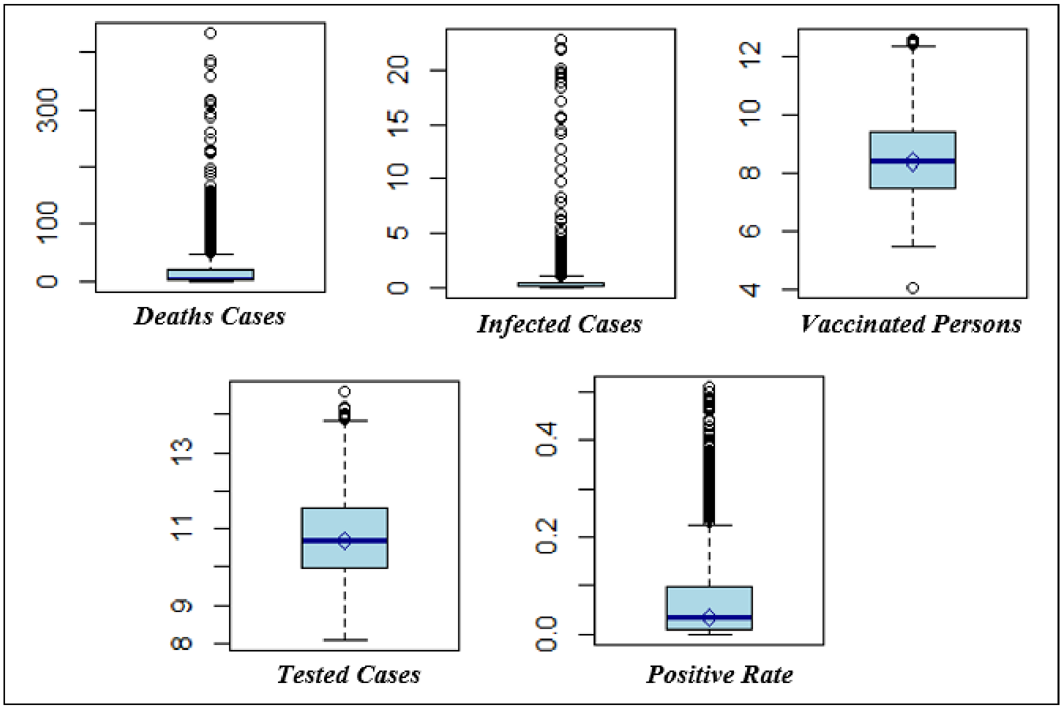

FIGURE 1. Box plot of the variables under study

significance level, where the null hypothesis was rejected which states that all panels contain unit roots, unlike the alternative hypothesis which states that panels are stationarity.

TABLE 4. Results of Panel Unit Root Test

	Deaths Cases	Infected Cases	Vaccinated Persons	Tested Cases	Positive Rate
Statistic	459.9	61.49	111	616.82	667.05
P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

4.4. Testing the Multicollinearity. The multicollinearity represents a serious problem when analyzing panel data, so it is important to try to confirm that there is not a high linear correlation among the independent variables, where multicollinearity is a type of disturbance that can occur in the data. When it is present in the data, the statistical conclusions and inferences about the dataset may not be reliable. This can lead to inaccurate estimates of the regression parameters, increased standard errors of these coefficients, inaccurate non-significant p-values, and reducing the predictive capability of the model, see [14, 40].

We can use a pairwise correlation matrix between independent variables and the variance inflation factor (VIF) to diagnose the linear relation between two or more of the explanatory variables as shown in Table 5.

TABLE 5. Results of Correlation Matrix and Variance Inflation Factor

Variables	Infected Cases	Vaccinated Persons	Tested Cases	Positive Rate
Infected Cases	1			
Vaccinated Persons	0.1471***	1		
Tested Cases	0.3455***	0.3430***	1	
Positive Rate	0.3062***	-0.3262***	-0.1444***	1
VIF	1.8050	1.1110	1.7035	1.3177

Note: The symbol *** indicates statistical significance at a level < 0.0001.

The results in Table 5 show that there are no high linear correlations among all regressors. If the correlation coefficients among independent variables exceed 0.8, the problem of multicollinearity appears in the data and becomes a serious problem. Also, the results of VIF in this table emphasized that there is no multicollinearity between the independent variables because all values of VIF were less than 5. Therefore, the multicollinearity is not a concern when analyzing this study. Where the usual rule of thumb in most econometrics studies is that any variable with a VIF exceeding 10 indicates the presence of a multicollinearity problem, see [12, 41, 42].

4.5. Fixed Effects Values of Cross-Sections.

Fig. 2 illustrates the graphic representation of the EF for all cross-sections in the study. Based on these results, it can be noted that the FE in Italy are 27.53, which is positive and high when compared to the other countries. This might be due to high infection rates. While the smallest value of the FE is for Denmark country, it is -10.89, which is low. The FE for the other countries range between the two previous values.

4.6. Redundant Fixed Effects Test.

To confirm the presence of the FE between the cross-sections, it was carried out the redundant FE test, and the results are shown in Table 6. The results of the test emphasize that the statistical values of the cross-section F and Chi-square are significant at < 0.0001 significance level, indicating that the FE across countries are different from one country to another.

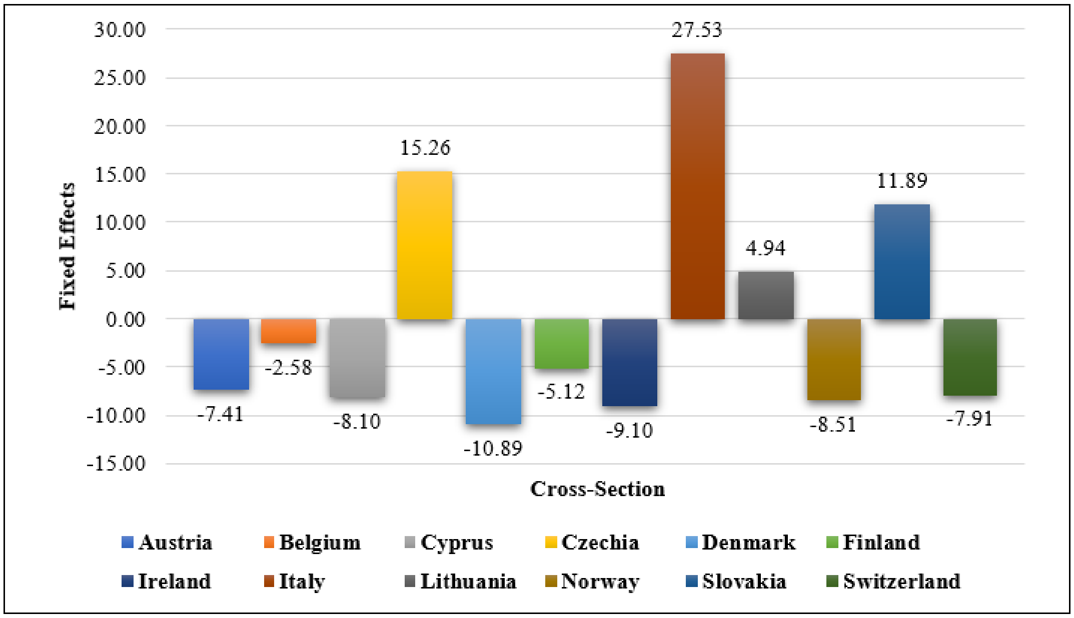

FIGURE 2. Fixed effect for the countries under study

TABLE 6. Results of Cross-Section Fixed Effects Test

Effects Test	Statistic	DF	P-value
Cross-section F	73.678155	(11, 2540)	< 0.0001
Cross-section Chi-square	707.841977	11	< 0.0001

4.7. Analysis of Variance Tests. The results displayed in Table 7 show that the value of the classical analysis of variance (ANOVA) F-test is equal to 105.240 and Welch's ANOVA F*-Test value is 119.134 with a p-value less than 0.0001, indicating that the ANOVA tests of equality of means are of high significance. This implies that the number of new deaths of COVID-19 are different from one country (cross-section) to another. For more information about the ANOVA tests, see e.g. [43].

TABLE 7. Analysis of Variance Tests of Equality of Means for Death Cases

Method	Test	Statistic	P-value
Classical ANOVA	F-Test	105.240	< 0.0001
Welch's ANOVA	F*-Test	119.134	< 0.0001

4.8. Results of Non-Robust Estimation Methods. Table 8 displays the results of non-robust estimation for FEP and FENB models. It can be seen that all the independent variables in the classical estimation methods have a statistically significant impact on the dependent variable, where infected cases, tested cases, and positive rates have a positive effect on death cases, but vaccinated persons have a negative impact.

TABLE 8. Coefficients Estimate for Classical Count Panel Data Models

	FEP		FENB	
Variables	Estimate	Std.E.	Estimate	Std.E.
Intercept	2.634***	0.006	2.544***	0.028
Infected Cases	0.127***	0.001	0.343***	0.020
Vaccinated Persons	-0.169***	0.005	-0.250***	0.029
Tested Cases	0.572***	0.011	0.310***	0.063
Positive Rate	1.532***	0.062	10.169***	0.491

Note: The symbol *** indicates statistical significance at a level < 0.0001.

4.9. Results of proposed Robust Estimation Methods. Table 9 displays the results of robust estimation techniques for the FEP model. We estimated the coefficients using an iteratively reweighted ML method. It can be noted that all the independent variables in the proposed estimation methods have a statistically significant impact on the dependent variable, so the estimated parameters are suitable for robust methods of the FEP model, where infected cases and vaccinated persons have a negative effect on deaths cases, but tested cases and positive rate have a positive impact.

4.10. Evaluating the Performance of Non-Robust and Proposed Robust Methods. Fig. 3 introduces the box plot for the residuals of FEP and FENB non-robust estimation methods. We can notice that the non-robust estimates of the FEP and FENB models are affected by the presence of outliers, as shown in Fig. 3. Therefore, we cannot use the negative binomial model with FE as an alternative to the Poisson model with FE to handle the problem of the presence of outliers to obtain robust estimates.

TABLE 9. Coefficients Estimate for Poisson Panel Data Model

	MFEP		SFEP		MMFEP	
Variables	Estimate	Std.E.	Estimate	Std.E.	Estimate	Std.E.
Intercept	1.575***	0.011	1.832***	0.013	1.912***	0.012
Infected Cases	-0.258***	0.024	-0.324***	0.024	-0.331***	0.022
Vaccinated Persons	-0.213***	0.009	-0.206***	0.011	-0.202***	0.010
Tested Cases	0.454***	0.025	0.530***	0.030	0.537***	0.028
Positive Rate	3.300***	0.155	3.659***	0.175	3.706***	0.164

Note: The symbol *** indicates statistical significance at a level < 0.0001.

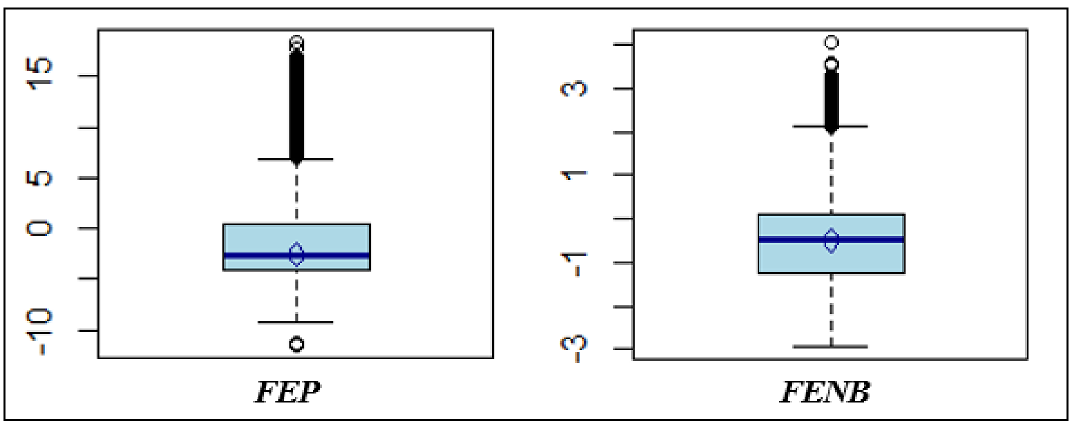

FIGURE 3. Disturbances of Non-Robust Estimation Methods

Fig. 4 shows the box plot for the residuals of MFEP, SFEP, and MMFEP robust estimation methods. We note that the performance of the proposed estimation methods is better than traditional methods for analyzing daily COVID-19 datasets.

To assess the performance of the proposed estimation methods, we used the mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), see, e.g. [44, 45, 17].

Based on the results of Table 10, we can examine the performance of the proposed estimation methods by applying them to the daily new deaths of COVID-19 cases by comparing the values of MSE, MAE, and MAPE.

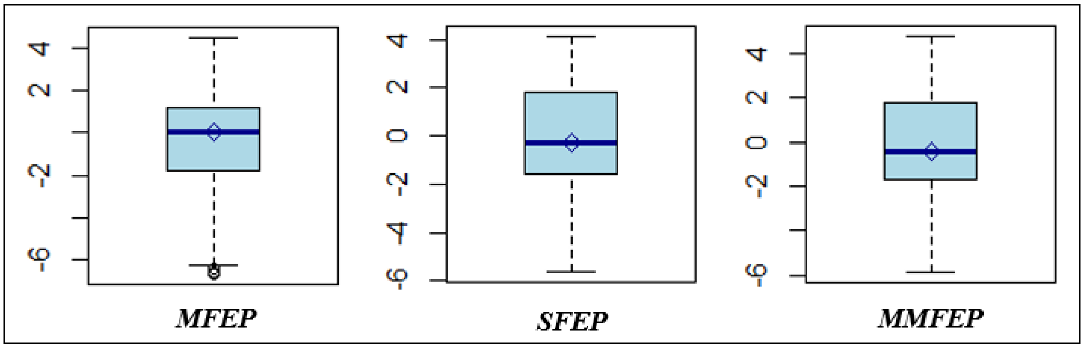

FIGURE 4. Disturbances of Robust Estimation Methods

TABLE 10. The Performance of Robust Methods

Robust Estimators	MSE	MAE	MAPE
MFEP	3.2917	2.3154	0.0464
SFEP	0.1388	0.5077	0.0057
MMFEP	0.0024	0.0657	0.0007

Table 10 indicates that the SFEP estimation method is better than the MFEP method, while the MMFEP method is better than the other proposed methods based on the values of MSE, MAE, and MAPE.

5. CONCLUSION

In this paper, we have proposed three robust (MFEP, SFEP, and MMFEP) estimation methods for the FEP model by applying daily COVID-19 datasets for 12 high-income countries in Europe during the period from June 23, 2021, to January 21, 2022. Moreover, our proposed new algorithms for robust Poisson with FE produce robust coefficient estimates and are useful for handling outliers. The results indicated that the estimates of non-robust FEP and FENB are highly sensitive to outliers, while proposed robust estimators are more efficient than FEP and FENB. In addition, the MMFEP estimation method is more efficient than SFEP and MFEP robust estimation methods because the MMFEP has minimum MSE, MAE, and MAPE values. In future research, we can develop a new robust estimation method for the FEP model in case of missing values and seasonality in the data as in [46].

Conflict of Interests

The authors declare that there is no conflict of interests.

References

[1] Y. Croissant, G. Millo, Panel data econometrics with R, First edition, John Wiley & Sons, Hoboken, 2019.

[2] C. Hsiao, Analysis of panel data, Cambridge University Press, Cambridge, 2014.

[3] W. Greene, Econometric analysis, Prentice Hall, Boston, 2012.

[4] B.H. Baltagi, Econometric analysis of panel data, J. Wiley & Sons, Chichester, 2005.

[5] J.F. Angers, D. Desjardins, G. Dionne, F. Guertin, Modelling and estimating individual and firm effects with count panel data, ASTIN Bull. 48 (2018), 1049–1078. https://doi.org/10.1017/asb.2018.19.

[6] S.S.P. Kumara, H.C. Chin, Study of fatal traffic accidents in Asia pacific countries, Transport. Res. Rec.: J. Transport. Res. Board 1897 (2004), 43–47. https://doi.org/10.3141/1897-06.

[7] R.B. Noland, M.A. Quddus, Analysis of pedestrian and bicycle casualties with regional panel data, Transport. Res. Rec.: J. Transport. Res. Board 1897 (2004), 28–33. https://doi.org/10.3141/1897-04.

[8] A.H. Youssef, M.R. Abonazel, E.G. Ahmed, The performance of count panel data estimators: a simulation study and application to patents in Arab countries, J. Math. Comput. Sci. 11 (2021), 8173–8196. https://doi.org/10.28919/jmcs/5852.

[9] A.C. Cameron, P.K. Trivedi, Regression analysis of count data, Cambridge University Press, Cambridge, 2013.

[10] J.P. Boucher, M. Denuit, Fixed versus random effects in Poisson regression models for claim counts: a case study with motor insurance, ASTIN Bull. 36 (2006), 285–301. https://10.2143/AST.36.1.2014153.

[11] J. Hausman, B. Hall, Z. Griliches, Econometric models for count data with an application to the patents-R&D relationship, National Bureau of Economic Research, Cambridge, MA, 1984. https://doi.org/10.3386/t0017.

[12] A.H. Youssef, M.R. Abonazel, E.G. Ahmed, Estimating the number of patents in the world using count panel data models, Asian J. Probab. Stat. 6 (2020), 24–33.

[13] M.R. Abonazel, I. Dawoud, Developing robust ridge estimators for Poisson regression model, Concurr. Comput.: Pract. Exper. 34 (2022), e6979. https://doi.org/10.1002/cpe.6979.

[14] I. Dawoud, M.R. Abonazel, Robust Dawoud–kibria estimator for handling multicollinearity and outliers in the linear regression model, J. Stat. Comput. Simul. 91 (2021), 3678–3692. https://doi.org/10.1080/00949655.2021.1945063.

[15] I.A. Idriss, W. Cheng, Robust estimators for Poisson regression, Open J. Stat. 13 (2023), 112–118. https://doi.org/10.4236/ojs.2023.131007.

[16] B.H. Beyaztas, S. Bandyopadhyay, Data driven robust estimation methods for fixed effects panel data models, J. Stat. Comput. Simul. 92 (2022) 1401–1425. https://doi.org/10.1080/00949655.2021.1996576.

[17] A Comparative Study of Robust Estimators for Poisson Regression Model with Outliers, J. Stat. Appl. Probab. 9 (2020), 279–286. https://doi.org/10.18576/jsap/090208.

[18] M. Amelia, K. Sadik, B. Sartono, The study of robust estimators on panel data regression model for data contaminated with outliers, in: Proceedings of the Proceedings of the 1st International Conference on Statistics and Analytics, 2-3 August 2019, Bogor, Indonesia, 2020. https://doi.org/10.4108/eai.2-8-2019.2290517.

[19] W. Bari, B.C. Sutradhar, Robust inferences in longitudinal models for binary and count panel data in the presence of outliers, Sankhya B 72 (2010), 11–37. https://doi.org/10.1007/s13571-010-0002-8.

[20] M.C. Bramati, C. Croux, Robust estimators for the fixed effects panel data model, Econometrics J. 10 (2007), 521–540. https://doi.org/10.1111/j.1368-423X.2007.00220.x.

[21] J.Á. Víšek, Estimating the model with fixed and random effects by a robust method, Methodol. Comput. Appl. Probab. 17 (2015), 999–1014. https://doi.org/10.1007/s11009-014-9432-5.

[22] Y. Susanti, H. Pratiwi, S. Sulistijowati H., T. Liana, M estimation, S estimation, and MM estimation in robust regression, Int. J. Pure Appl. Math. 91 (2014), 349–360. https://doi.org/10.12732/ijpam.v91i3.7.

[23] P. Huber, Robust estimation of a location parameter, Ann. Math. Stat. 35 (1964), 73–101.

[24] F.M. Alghamdi, A.R. Kamel, M.S. Mustafa, et al. A statistical study for the impact of REMS and nuclear energy on carbon dioxide emissions reductions in G20 countries, J. Rad. Res. Appl. Sci. 17 (2024), 100993. https://doi.org/10.1016/j.jrras.2024.100993.

[25] A.H. Youssef, M.R. Abonazel, A.R. Kamel, Efficiency comparisons of robust and non-robust estimators for seemingly unrelated regressions model, WSEAS Trans. Math. 21 (2022), 218–244. https://doi.org/10.37394/23206.2022.21.28.

[26] N. Wang, Y.G. Wang, S. Hu, et al. Robust regression with data-dependent regularization parameters and autoregressive temporal correlations, Environ. Model. Assess. 23 (2018), 779–786. https://doi.org/10.1007/s10666-018-9605-7.

[27] O. Toka, M. Cetin, The comparing of S-estimator and M-estimators in linear regression, Gazi Univ. J. Sci. 24 (2011), 747–752.

[28] O. Arslan, O. Edlund, H. Ekblom, Algorithms to compute CM- and S-estimates for regression, Metrika 55 (2002), 37–51. https://doi.org/10.1007/s001840200185.

[29] P. Rousseeuw, V. Yohai, Robust regression by means of S-estimators, in: J. Franke, W. Härdle, D. Martin (Eds.), Robust and Nonlinear Time Series Analysis, Springer, New York, 1984: pp. 256–272. https://doi.org/10.1007/978-1-4615-7821-5_15.

[30] B.H. Baltagi, The Oxford handbook of panel data, Oxford University Press, 2015.

[31] V.J. Yohai, High breakdown-point and high efficiency robust estimates for regression, Ann. Stat. 15 (1987), 642–656. https://www.jstor.org/stable/2241331.

[32] M.R. Abonazel, N.M. Darwish, Forecasting confirmed and recovered COVID-19 cases and deaths in Egypt after the genetic mutation of the virus: ARIMA Box-Jenkins approach, Commun. Math. Biol. Neurosci. 2022 (2022), 17. https://doi.org/10.28919/cmbn/6888.

[33] R.M. Ghazy, A. Gebreal, B.E. El Demerdash, et al. Development and validation of a French questionnaire that assesses knowledge, attitude, and practices toward Marburg diseases in sub-Saharan African countries, Public Health 230 (2024), 128–137. https://doi.org/10.1016/j.puhe.2024.01.027.

[34] R. Arunachalam, T. Pakkirisamy, Panel data modelling of COVID-19 infected cases, preprint, (2021). https://doi.org/10.21203/rs.3.rs-635014/v1.

[35] H. Çivak, C. Küren, S. Turgay, Examining the effects of COVID-19 data with panel data analysis, Soc. Med. Health Manage.2 (2021), 1–16.

[36] N. Thomas, A. Jana, S. Bandyopadhyay, Effect of policies and COVID-19 cases on mobility and activity participation during 300 days of COVID-19: panel data analysis of Asian countries, J. Eastern Asia Soc. Transport. Stud. 14 (2022), 16–33. https://doi.org/10.11175/easts.14.16.

[37] M.R. Abonazel, A.R. Rabie, The impact of using robust estimations in regression models: An application on the Egyptian economy, J. Adv. Res. Appl. Mat. 4 (2019), 8–16.

[38] Ö.G. Alma, Comparison of robust regression methods in linear regression, Int. J. Contemp. Math. Sci. 6 (2011), 409–421.

[39] M.R. Abonazel, O.A. Shalaby, Using dynamic panel data modeling to study net FDI inflows in MENA countries, Stud. Econ. Econometrics 44 (2020), 1–28.

[40] R.K. Paul, Multicollinearity: causes, effects and remedies, Indian Agricultural Statistics Research Institute, New Delhi, 2006.

[41] F. Sun, S.A. Matthews, T.C. Yang, et al. A spatial analysis of the COVID-19 period prevalence in u.s. counties through June 28, 2020: where geography matters?, Ann. Epidemiol. 52 (2020), 54-59.e1. https://doi.org/10.1016/j.annepidem.2020.07.014.

[42] D.N. Gujarati, D.C. Porter, Basic econometrics, Tata McGraw-Hill Education, 2009.

[43] M. Delacre, C. Leys, Y.L. Mora, et al. Taking parametric assumptions seriously: arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA, International Review of Social Psychology 32 (2019), 13. https://doi.org/10.5334/irsp.198.

[44] A.A.K.A. Hamid, W.I.A.W.M. Nawi, M.S. Lola, et al. Improvement of time forecasting models using machine learning for future pandemic applications based on covid-19 data 2020–2022, Diagnostics 13 (2023), 1121. https://doi.org/10.3390/diagnostics13061121.

[45] W.I.A.W.M. Nawi, M.S. Lola, R. Zakariya, et al. Improved of forecasting sea surface temperature based on hybrid arima and support vector machines models, Malays. J. Fundam. Appl. Sci. 17 (2021), 609–620. https://doi.org/10.11113/mjfas.v17n5.2356.

[46] U. Islam, B. Shah, A.A. Al-Atawi, et al. Empowering global ethereum price prediction with EtherVoyant: a state-of-the-art time series forecasting model, Neural Comput. Appl. (2024). https://doi.org/10.1007/s00521-024-10169-3.