frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 26 August 2025
DOI 10.3389/fams.2025.1638596

Check for updates

OPEN ACCESS

EDITED BY
Coskun Kus,
Selcuk University, Türkiye

REVIEWED BY
Zakariya Yahya Algamal,
University of Mosul, Iraq
Özlem Ege Oruç,
Dokuz Eylul University, Türkiye
Caner Tanış,
Cankiri Karatekin University, Türkiye

*CORRESPONDENCE
Mohamed R. Abonazel
mabonazel@cu.edu.eg

RECEIVED 31 May 2025
ACCEPTED 04 August 2025
PUBLISHED 26 August 2025

CITATION
Abonazel MR, Ebrahim EEM, Ahmed EG and El-Masry AM (2025) New robust estimators for the fixed effects negative binomial model: a simulation and real-world applications to European panel data.
*Front. Appl. Math. Stat.* 11:1638596.
doi: 10.3389/fams.2025.1638596



# New robust estimators for the fixed effects negative binomial model: a simulation and real-world applications to European panel data

Mohamed R. Abonazel[1]*, Ehab Ebrahim Mohamed Ebrahim[2], Elsayed G. Ahmed[3] and Amera M. El-Masry[4]

[1]Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt, [2]Department of Economics, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, [3]Department of Statistics and Insurance, Faculty of Commerce, Arish University, Arish, Egypt, [4]Department of Mathematics and Statistics, Faculty of Management Technology and Information Systems, Port Said University, Port Said, Egypt

This study introduces a more flexible approach by employing the fixed effects negative binomial model to address challenges associated with outliers and dispersion. Unlike previous studies that focused on the robust estimation of the Poisson model with fixed effects, which assumes equidispersion and cannot handle dispersion in count panel data, we develop novel estimators specifically designed for the fixed effects negative binomial panel regression model in the presence of outliers, under-dispersion, and over-dispersion. The methodology is assessed through comprehensive simulation experiments across different scenarios. A comprehensive empirical analysis is conducted using updated and extended panel datasets on COVID-19 and patent applications in Europe. The results of both Monte Carlo simulation and the empirical studies indicate that the robust estimators: the robust fixed negative binomial Huber, fixed negative binomial Hampel, and fixed negative binomial Tukey estimators, outperform the classical non-robust fixed negative binomial estimator.

KEYWORDS

**robust estimation, outliers, dispersion, count panel data, fixed effects negative binomial model, conditional maximum likelihood, *M* estimation**

## 1 Introduction

The negative binomial (NB) regression model has become a widely utilized tool in panel data analysis across various fields, including economics, healthcare, demography, and environmental sciences (see Ahmed et al. [1], Abonazel et al. [2], and Youssef et al. [3]). Panel data analysis has emerged as a prominent area of research in econometrics literature as it involves studying data along both the temporal and sectoral dimensions to derive the maximum benefit from the available data. Panel datasets combine cross-sectional and time-series observations, enabling a more in-depth analysis of the data over time [4, 5]. By including individual-specific fixed effects, panel data can account for these differences and provide more accurate estimates of the effects of the variables of interest. Within this framework, count panel data models are particularly relevant when the dependent variable represents non-negative integer values generated from counting the occurrence of a particular event, for example, the number of accidents on roads, patents granted to countries and companies, and the number of deaths due to a specific disease (see Winkelmann [6]).

One of the key challenges in analyzing count panel data is the issue of over-dispersion; this problem appears when the variance of the dependent variable is greater than the mean. Therefore, the NB panel model has a dispersion parameter $\phi_i$; this parameter allows $\text{var}(y_{it}) > \text{E}(y_{it})$. This violates the core assumption of the classical Poisson model and often leads to inefficient or biased estimates. To address this limitation, the NB model is a powerful statistical model that is commonly used in analyzing count data. Unlike the Poisson model, the NB model is capable of handling over-dispersion in count panel data; this characteristic makes it an important model in various fields, such as economics, finance, public health, criminology, environmental health, transportation, finance, and law, where it provides a more flexible and accurate approach for analyzing count data. Therefore, several studies have focused on the use of the NB model. For example, Cameron and Trivedi [7], Yaacob et al. [8], Hana [9], Al-Taweel and Algamal [10], Alobaidi et al. [11], Rashad et al. [12], Niang et al. [13], Liu et al. [14], and Suryadi et al. [15].

Therefore, the FENB model is great importance in statistical analysis due to its ability to address over-dispersion in count panel data and control unobserved heterogeneity among individuals or entities in panel data. The model captures individual-specific or entity-specific characteristics that are constant over time. The FENB model finds applications in various fields such as economics, public health, criminology, and social sciences. Youssef et al. [16] conducted a study on estimating the number of patents in some high-income countries using the NB model and Poisson model based on the panel data models. The research aimed to develop accurate models for predicting patents. The results indicate that the FENB model is more suitable and performs better in analyzing the data. This research offers valuable contributions to the field of patent analysis by providing reliable models for estimating patents, enhancing our understanding of the factors that influence global patent trends, and providing useful insights for researchers and policymakers in the field of estimating patents. Kumara and Chin [17] conducted a study analyzing accident data for 41 countries over 15 years. They used the FENB model to examine the effect of socioeconomic and infrastructure factors on fatal road accidents in the Asia-Pacific region. The study also employed the AIC statistic to assess the suitability of the model. Guimaraes [18] used the FENB model under a specific set of assumptions. The study showed that this model effectively addresses over-dispersion. This research emphasizes the significance of the FENB model for accurately analyzing count panel data, particularly in the presence of over-dispersion. In the FENB model, Hausman et al. [19] added the individual fixed effects $\alpha_i$ and the NB dispersion parameter ($\phi_i$) into the model. Then, the FENB model given is

$$f(y_{it}|\theta_i) = \frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)}\left(\frac{1}{1+\theta_i}\right)^{\lambda_{it}}\left(\frac{\theta_i}{1+\theta_i}\right)^{y_{it}}, \quad i = 1,2,\ldots,N; \quad t = 1,2,\ldots,T, \tag{1}$$

where $y_{it}$ is the dependent variable and takes values of non-negative integers, i.e., $y_{it} \geq 0$. The mean and the variance of the NB model with fixed effects are not equal. $\Gamma(\cdot)$ is the gamma function, and $\theta_i = \frac{\alpha_i}{\phi_i}$ is the parameter of the individual effects and dispersion, which is assumed to be constant over time for each cross section, while $\lambda_{it}$ depends on covariates by the function $\ln(\lambda_{it}) = X'_{k,it}\beta$.

We can use the conditional maximum likelihood (CML) estimation method to estimate the parameters of the FENB model [19]. The NB model with fixed effects assumes that for a given cross section ($i$), the dependent variable ($y_{it}$) is independent over time and $\sum_{t=1}^{T} y_{it}$ has a NB distribution with parameters $\theta_i$ and $\sum_{t=1}^{T}\lambda_{it}$, where

$$\sum_{t=1}^{T} y_{it} \sim \text{NB}\left[\theta_i\sum_{t=1}^{T}\lambda_{it}, \left(\theta_i\sum_{t=1}^{T}\lambda_{it}\right)(1+\theta_i)\right].$$

For an individual ($i$), the dependent variable $y_{it}$ is independent over time as follows:

$$f\left(\sum_{t=1}^{T} y_{it}\right) = \left[\frac{\Gamma\left(\sum_{t=1}^{T}\lambda_{it}+\sum_{t=1}^{T}y_{it}\right)}{\Gamma\left(\sum_{t=1}^{T}\lambda_{it}\right)\Gamma\left(\sum_{t=1}^{T}y_{it}+1\right)}\right]\left[\left(\frac{1}{1+\theta_i}\right)^{\sum_{t=1}^{T}\lambda_{it}}\right]\left[\left(\frac{\theta_i}{1+\theta_i}\right)^{\sum_{t=1}^{T}y_{it}}\right].$$

The joint probability function for the $i^{th}$ observation for Equation 1 is

$$f(y_{i1},\ldots,y_{iT}) = \prod_{t=1}^{T}\left\{\left[\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)}\right]\left[\left(\frac{1}{1+\theta_i}\right)^{\lambda_{it}}\right]\left[\left(\frac{\theta_i}{1+\theta_i}\right)^{y_{it}}\right]\right.$$

Then, the conditional joint probability function for the $i^{th}$ observation is

$$f(y_{i1},\ldots,y_{iT}|\sum_{t=1}^{T}y_{it}) = \left[\frac{\Gamma\left(\sum_{t=1}^{T}\lambda_{it}\right)\Gamma\left(\sum_{t=1}^{T}y_{it}+1\right)}{\Gamma\left(\sum_{t=1}^{T}\lambda_{it}+\sum_{t=1}^{T}y_{it}\right)}\right]. \times\left(\prod_{t=1}^{T}\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)}\right)$$

The CML estimation of the NB model with fixed effects can be obtained by maximizing the following log CML function:

$$L(y_{i1},\ldots,y_{iT}\mid\sum_{t=1}^{T}y_{it}) = \sum_{i=1}^{N}\left\{\ln\Gamma\left(\sum_{t=1}^{T}\lambda_{it}\right)+\ln\Gamma\left(\sum_{t=1}^{T}y_{it}+1\right)\right. - \ln\Gamma\left(\sum_{t=1}^{T}\lambda_{it}+\sum_{t=1}^{T}y_{it}\right)+\sum_{t=1}^{T}\left[\ln\Gamma(\lambda_{it}+y_{it})-\ln\Gamma(\lambda_{it})\right. \left. - \ln\Gamma(y_{it}+1)\right]\Big\}.$$

For a more comprehensive understanding of CML estimation and its applications see Yousse et al. [16, 20]. In the context, the

FENB model is particularly useful because it not only accounts for over-dispersion but also controls for unobserved, time-invariant heterogeneity among individuals. This makes the FENB model a powerful tool for analyzing panel count data with complex structures. However, one major issue that has been neglected in the literature is the effect of outliers on parameters estimation in the FENB model. Outliers can occur due to rare events or abnormal changes in specific units over time. Classical estimation techniques, such as the CML, are known to be highly sensitive to such outliers. This highlights the need for robust estimation methods that can produce accurate and stable results even in the presence of outliers. Despite the progress made in developing robust estimators for the fixed effects Poisson model by Ahmed et al. [1], Abonazel et al. [2], and Youssef et al. [3], there has been little interest in extending these robust methods to the NB model. This gap in literature provides a key motivation for the present study.

## 2 Robust FENB model estimates

According to the work of Youssef et al. [3], who introduced robust estimation methods for the Poisson fixed effects panel model, this study extends the robustness framework to the NB case to accommodate both under-dispersion and over-dispersion in count panel data. Accordingly, this section aims to develop robust estimators for the FENB model in the presence of dispersion and outliers, based on the Huber, Hampel, and Tukey weight functions. The performance of these estimators is evaluated through Monte Carlo simulations under varying levels of outliers, dispersion levels, cross-sectional sizes, and time dimensions. Furthermore, the proposed methods are applied to real-world datasets in two empirical applications involving selected European countries, providing strong evidence of their practical effectiveness.

The NB model addresses over-dispersion that may exist in the data, where the model provides more accurate and reliable estimates; this makes it suitable for a wide range of applications in various fields. The count panel data can contain outliers, and these outliers can have a significant impact on the estimated parameters, where some studies focus on addressing outliers in some models using different estimation methods (see for example Huber [21, 22], Rousseeuw and Leroy [23], Garay et al. [24], Toka and Cetin [25], Susanti et al. [26], Maronna et al. [27], Tüzun et al. [28], Youssef et al. [3, 29], and Ke et al. [30]). On the other hand, panel data regression models may suffer from outliers that can negatively affect classical estimation methods, where there are limited studies on robust estimation methods for fixed effects panel data regression model (see Beyaztas and Bandyopadhyay [31], Amelia et al. [32], Víšek [33], and Bramati and Croux [34]).

In this context, there is not a specific robust method available for estimating parameters in the FENB regression model. Therefore, we propose a robust estimation method for NB regression to deal with count panel data when these data contain outliers. The method utilizes robust estimation techniques that are less sensitive to the influence of outliers, providing more accurate and robust parameter estimates. These methods are based on weight functions that include Huber, Hampel, and Tukey bisquare and minimize the residuals, aiming to improve the robustness and reliability of parameter estimation in the presence of outliers, see for example Hampel et al. [35] and Venables and Ripley [36]. The residuals of the $i^{\text{th}}$ observation for the FENB model can be expressed as

$$R_{it} = y_{it} - \left( \frac{\alpha_i}{\phi_i} \sum_{t=1}^{T} \exp(X'_{k,it}\beta) \right).$$

The robust estimator for the FENB model can be obtained by minimizing the objective function ($\rho$) concerning all $\beta$, which can be expressed as follows:

$$\hat{\beta}_R = \min \sum_{i=1}^{N} \sum_{t=1}^{T} \rho \left( \frac{\xi_{it}}{\hat{\sigma}_{MAD}} \right), \tag{2}$$

where $\xi_{it} = \frac{R_{it}}{\left(\hat{\theta}_i \sum_{t=1}^{T} \lambda_{it}\right)^{1/2}}$, $\lambda_{it} = \exp(X'_{k,it}\beta)$, and $\hat{\sigma}_{MAD}$ is the estimated standard deviation, where calculated using the median absolute deviation, which is defined as follows:

$$\hat{\sigma}_{MAD} = 1.4826 \times Med\left(|\xi_{it} - Med(\xi_{it})|\right),$$

where $Med$ is the median. Based on the function $\rho(\xi_{it})$, we can obtain the robust estimator of $\beta$ by differentiating the objective function ($\rho$) concerning the parameters $\beta$, and equate the partial derivatives to zero in Equation 2, this yields the following expression:

$$\min_{\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} \psi \left( \frac{\xi_{it}}{\hat{\sigma}_{MAD}} \right) \frac{\partial}{\partial \beta} \xi_{it} = 0, \tag{3}$$

where $\psi(u_{it})$ is called the influence function and equal $\rho'(u_{it})$. We can rewrite the Equation 3 by using the weight function $_{\xi}(u_{it})$, which equals $\frac{\psi(u_{it})}{u_{it}}$, and then, we can obtain the first-order condition for robust estimators as follows:

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \xi'_{it} W_{\xi}(\xi_{it}) \xi_{it} = 0, \tag{4}$$

where $\xi'_{it} = \frac{\partial}{\partial \beta} \xi_{it}$ and $W_{\xi}(\xi_{it})$ is matrix $(NT \times NT)$. Three robust estimates for $\beta$ in the FENB model can be obtained by solving Equation 4 using the weight functions of Huber ($W_{HR}(\xi_{it})$), Hampel ($W_{HM}(\xi_{it})$), and Tukey bisquare ($W_{TK}(\xi_{it})$), as introduced by Baltagi [37] and Youssef et al. [3] in the context of panel data.

We will present an algorithm for handling outliers in the NB panel data model with fixed effects to obtain stable and robust estimates as follows:

1. Estimate the CML estimator ($\hat{\beta}_{\text{NOFNB}}$) using the algorithm presented by Hausman et al. [19] as an initial estimation.
2. Calculate residuals value $\xi_{it} = \frac{R_{it}}{\left(\hat{\theta}_i \sum_{t=1}^{T} \lambda_{it}\right)^{1/2}}$ based on the previous estimation.
3. Calculate median absolute deviation $\hat{\sigma}_{\text{MAD}}$.
4. Calculate standardized residuals $u_{it} = \frac{\xi_{it}}{\hat{\sigma}_{\text{MAD}}}$.
5. Calculate the values of the weight function $W_{\xi}(\xi_{it})$. We can use one of the three functions: $W_{HR}(\xi_{it})$, $W_{HM}(\xi_{it})$, and $W_{TK}(\xi_{it})$.
6. Estimate one of the three estimators ($\hat{\beta}_{\text{FNBHR}}, \hat{\beta}_{\text{FNBHM}}$, or $\hat{\beta}_{\text{FNBTK}}$) based on the weight function $W_{\xi}(\xi_{it})$ used in the previous step.

7. Repeat steps 2–6 to obtain the convergent estimates of $\hat{\beta}_{\text{FNBHR}}, \hat{\beta}_{\text{FNBHM}}$, or $\hat{\beta}_{\text{FNBTK}}$ estimators.
8. Examine the significance of the explanatory variables on the response variable and compare the performance of these estimators using some criteria; for example, the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The AIC and BIC are calculated as follows:

$$\text{AIC} = 2k - 2\log(likhod); \quad \text{BIC} = k\log(n) - 2\log(likhod),$$

where $k$ is the number of coefficients, and *likhod* denotes maximum likelihood values.

# 3 Simulation design and evaluation of estimators

In this section, we will present an algorithm to evaluate the impact of outliers on the estimates of the FENB model using a Monte Carlo simulation procedure to conduct a comprehensive analysis to understand the flexibility of the NB model with fixed effects in the presence of outliers. The simulation was conducted using R software. This algorithm was designed specifically for the Monte Carlo simulation study of the FENB model, where it aims to assess the model's sensitivity and accuracy in various conditions. The results from this simulation are expected to contribute significantly to the existing literature, offering an accurate perspective on the applicability and reliability of the FENB model in empirical research, especially in fields where extreme values are not merely outliers but pivotal points of interest. The algorithm utilized for conducting the Monte Carlo simulation study of the FENB model is as follows:

1. We can construct the panel dataset to calculate the total samples size ($n$) as follows:
   (a) Set the dimensions for the cross-sectional size at $N = 50, 100, 200,$ and 400 to indicate small, moderate, and large units, respectively.
   (b) Set the dimensions for time-series size at $T = 5, 10,$ and 20 to indicate various sizes for the time periods.
2. We generate the count panel dataset by following these steps:
   (a) The values of true parameters for $\beta_1, \beta_2$, and $\beta_3$ were selected to be 1.
   (b) The dispersion parameter ($\phi$) was set at 1/2 and 2 to denote under-dispersion and over-dispersion, respectively.
   (c) The explanatory variables ($X_{it}^k$) were generated from a uniform distribution on the interval $(-1, 1)$.
   (d) Generate the response variable ($y_{it}$) from the NB distribution with an average $\theta_i \sum_{t=1}^{T} \lambda_{it}$.
   (e) Set the outliers ($\tau$%) to be 0%, 5%, 10%, 15%, and 20% of the total observations for the response variable. When the proportion of outliers is zero ($\tau = 0\%$), it indicates that the count panel dataset does not contain any outliers.
   (f) Generate the outliers from NB distribution with an average $4\,\text{IQR}[\theta_i \sum_{t=1}^{T} \lambda_{it}]$, where IQR is the interquartile range.
3. Estimate the non-robust estimator ($\hat{\beta}_{\text{NOFNB}}$) of the generated FENB model.
4. Estimate the proposed robust estimators ($\hat{\beta}_{\text{FNBHR}}, \hat{\beta}_{\text{FNBHM}}$, or $\hat{\beta}_{\text{FNBTK}}$) of the generated FENB model with the following steps:
   (a) Calculate residuals value $\xi_{it} = \frac{R_{it}}{\left(\hat{\theta}_i \sum_{t=1}^{T} \lambda_{it}\right)^{1/2}}$ based on the previous estimation.
   (b) Calculate median absolute deviation $\hat{\sigma}_{\text{MAD}}$.
   (c) Calculate standardized residuals $u_{it} = \frac{\xi_{it}}{\hat{\sigma}_{\text{MAD}}}$.
   (d) Calculate the values of the three functions: $W_{HR}(\xi_{it})$, $W_{HM}(\xi_{it})$, and $W_{TK}(\xi_{it})$.
   (e) Estimate the three robust estimators ($\hat{\beta}_{\text{FNBHR}}, \hat{\beta}_{\text{FNBHM}}$, and $\hat{\beta}_{\text{FNBTK}}$) based on the three weight functions $(W_{HR}(\xi_{it}), W_{HM}(\xi_{it}), and\ W_{TK}(\xi_{it}))$ calculated in the previous step.
   (f) Repeat steps (a)–(e) to obtain the convergent estimates of $\hat{\beta}_{\text{FNBHR}}, \hat{\beta}_{\text{FNBHM}}$, and $\hat{\beta}_{\text{FNBTK}}$ estimators.
5. For all simulation experiments, we conducted 500 replications.
6. We evaluate the performance of these estimators through the following steps:
   (a) The mean squared error (MSE) and mean absolute error (MAE) are calculated for $N, T, \phi$, and $\tau$ different for each parameter separately as follows:

$$\text{MSE} = \frac{1}{L}\sum_{l=1}^{L}(\hat{\beta}_l - \beta)^2; \quad \text{MAE} = \frac{1}{L}\sum_{l=1}^{L}|\hat{\beta}_l - \beta|,$$

   where $\hat{\beta}_l$ represents the estimated coefficient vector in the $l^{th}$ experiment of $L$ replicated simulations, and $\beta$ denotes the true coefficient vector.
   (b) The mean relative efficiency (MRE) is calculated for all $T$ and $\tau$ for each $N$ and $\phi$ separately to evaluate the efficiency of estimators (FNBHR, FNBHM, and FNBTK). The MRE is calculated using the following formula:

$$\text{MRE}(\hat{\beta}_R) = \frac{\text{MSE}(\hat{\beta}_{NOFNB})}{\text{MSE}(\hat{\beta}_R)},$$

   where $\hat{\beta}_R$ is $\hat{\beta}_{FNBHR}, \hat{\beta}_{FNBHM}$, and $\hat{\beta}_{FNBTK}$ estimators. The best efficient estimator is the one with the highest MRE.

The simulation results are presented in Tables 1–4 for small, moderate, and large sample sizes. These tables report the total MSE and total MAE values for four estimators: the classical (non-robust) estimator NOFNB and three robust estimators (FNBHM, FNBHR, and FNBTK). The simulation is conducted under various scenarios defined by cross-sectional sizes ($N = 50, 100, 200,$ and 400), time periods ($T = 5, 10,$ and 20), outlier proportions ($\tau = 0\%, 5\%, 10\%, 15\%,$ and 20%), and two dispersion levels representing under-dispersion ($\phi = 0.5$) and over-dispersion ($\phi = 2$).

In the context of clean data free from outliers ($\tau = 0\%$), the NOFNB estimator consistently performs better than the proposed estimators (FNBHM, FNBHR, and FNBTK) across all cases of $N$, $T$, and $\phi$. This suggests that under ideal data conditions, the classical CML method offers higher efficiency. For example, at $N = 50$, $\phi = 0.5$, $T = 5$, and $\tau = 0\%$, the MSE of the NOFNB estimator is 0.1614, whereas the values for FNBHM, FNBHR, and FNBTK are 0.2331, 0.2758, and 0.3345, respectively.

TABLE 1 MSE and MAE values of non-robust and robust estimates when $N$ = 50.

| Estimators | $\phi = 0.5$ | | | | | | $\phi = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| $\tau = 0\%$ | | | | | | | | | | | | |
| NOFNB | 0.1614 | 0.5515 | 0.0664 | 0.3523 | 0.0314 | 0.2446 | 0.0762 | 0.3793 | 0.0315 | 0.2464 | 0.0142 | 0.1656 |
| FNBHM | 0.2331 | 0.6614 | 0.0955 | 0.4261 | 0.0436 | 0.2872 | 0.0976 | 0.4322 | 0.0396 | 0.2733 | 0.0188 | 0.1900 |
| FNBHR | 0.2758 | 0.7072 | 0.1348 | 0.5081 | 0.0816 | 0.4048 | 0.1241 | 0.4884 | 0.0585 | 0.3347 | 0.0344 | 0.2629 |
| FNBTK | 0.3345 | 0.7700 | 0.1685 | 0.5677 | 0.0995 | 0.4456 | 0.1335 | 0.5048 | 0.0618 | 0.3433 | 0.0355 | 0.2656 |
| $\tau = 5\%$ | | | | | | | | | | | | |
| NOFNB | 0.3727 | 0.8426 | 0.2971 | 0.7767 | 0.2073 | 0.6861 | 0.2564 | 0.7132 | 0.1871 | 0.6603 | 0.1566 | 0.6306 |
| FNBHM | 0.2588 | 0.7021 | 0.1371 | 0.5184 | 0.0915 | 0.4344 | 0.0975 | 0.4323 | 0.0508 | 0.3135 | 0.0292 | 0.2441 |
| FNBHR | 0.2640 | 0.7070 | 0.1127 | 0.4606 | 0.0528 | 0.3150 | 0.1050 | 0.4433 | 0.0477 | 0.3015 | 0.0241 | 0.2181 |
| FNBTK | 0.3132 | 0.7710 | 0.1415 | 0.5179 | 0.0685 | 0.3594 | 0.1090 | 0.4509 | 0.0507 | 0.3107 | 0.0264 | 0.2282 |
| $\tau = 10\%$ | | | | | | | | | | | | |
| NOFNB | 0.6129 | 1.1273 | 0.5432 | 1.1247 | 0.4771 | 1.1143 | 0.5437 | 1.1225 | 0.5100 | 1.1533 | 0.4186 | 1.0810 |
| FNBHM | 0.3129 | 0.7806 | 0.2118 | 0.6632 | 0.1543 | 0.5980 | 0.1465 | 0.5298 | 0.0896 | 0.4330 | 0.0713 | 0.4049 |
| FNBHR | 0.3004 | 0.7551 | 0.1821 | 0.5954 | 0.1112 | 0.4769 | 0.1651 | 0.5624 | 0.1039 | 0.4664 | 0.0749 | 0.4118 |
| FNBTK | 0.3479 | 0.8151 | 0.2061 | 0.6397 | 0.1310 | 0.5193 | 0.1476 | 0.5292 | 0.0883 | 0.4242 | 0.0656 | 0.3800 |
| $\tau = 15\%$ | | | | | | | | | | | | |
| NOFNB | 0.9355 | 1.4571 | 0.7184 | 1.3505 | 0.7201 | 1.4062 | 0.7857 | 1.4148 | 0.7836 | 1.4760 | 0.6814 | 1.4033 |
| FNBHM | 0.4147 | 0.9153 | 0.2729 | 0.7778 | 0.2356 | 0.7659 | 0.2388 | 0.7199 | 0.1847 | 0.6628 | 0.1581 | 0.6456 |
| FNBHR | 0.4321 | 0.9243 | 0.2583 | 0.7342 | 0.2169 | 0.7118 | 0.2645 | 0.7587 | 0.2209 | 0.7292 | 0.1762 | 0.6820 |
| FNBTK | 0.4489 | 0.9383 | 0.2624 | 0.7399 | 0.2203 | 0.7162 | 0.2048 | 0.6546 | 0.1626 | 0.6126 | 0.1331 | 0.5837 |
| $\tau = 20\%$ | | | | | | | | | | | | |
| NOFNB | 1.1380 | 1.6473 | 1.0699 | 1.6821 | 0.9799 | 1.6663 | 1.0817 | 1.7015 | 1.0457 | 1.7270 | 0.9210 | 1.6420 |
| FNBHM | 0.5310 | 1.0483 | 0.3832 | 0.9432 | 0.3456 | 0.9515 | 0.3886 | 0.9513 | 0.3305 | 0.9201 | 0.2940 | 0.9068 |
| FNBHR | 0.5651 | 1.0733 | 0.4355 | 0.9947 | 0.3714 | 0.9754 | 0.4283 | 1.0034 | 0.3747 | 0.9865 | 0.3163 | 0.9405 |
| FNBTK | 0.5502 | 1.0441 | 0.3924 | 0.9313 | 0.3372 | 0.9221 | 0.3157 | 0.8375 | 0.2599 | 0.8015 | 0.2294 | 0.7927 |

In contrast, when the data contain outliers ($\tau > 0\%$), the FNBHM, FNBHR, and FNBTK estimators outperform the classical NOFNB estimator. The presence of outliers negatively affects the performance of NOFNB, leading to inflated error values, while the robust methods maintain greater stability and accuracy. For instance, at $N = 200$, $\phi = 2$, $T = 20$, and $\tau = 10\%$, the MSE values of FNBHM, FNBHR, and FNBTK are 0.0597, 0.0663, and 0.0566, respectively, compared to 0.4512 for NOFNB.

Moreover, as the proportion of outliers increases, both MSE and MAE tend to rise; however, the rate of increase is slower for the proposed estimators than for the NOFNB. For example, when $\tau$ increases from 10 to 15% at $N = 100$, $\phi = 2$, and $T = 10$, the MAE values of FNBHM, FNBHR, and FNBTK increase from 0.4100, 0.4087, and 0.3826 to 0.6216, 0.6759, and 0.5661, respectively, while the MAE of NOFNB jumps from 1.0490 to 1.4436. These results highlight the greater robustness of the proposed estimators in the presence of outliers.

As the number of cross-sectional units increases from $N = 50$ to $N = 400$, the values of MSE and MAE generally decrease for all estimators, indicating enhanced estimation accuracy due to the larger unit's size and increased information. For example, when $T = 10$, $\tau = 5\%$, and $\phi = 2$, the MSE of the classical estimator NOFNB decreases from 0.1871 at $N = 50$ to 0.1565 at $N = 400$. Similarly, for the proposed robust estimator FNBTK, the MSE drops from 0.0507 to 0.0123 over the same increase in $N$. This decreasing trend is consistently observed for both robust and non-robust estimators. However, robust estimators continue to offer more reliable estimates in the presence of outliers, regardless of the value of $N$. Although the classical estimator benefits significantly from increased sample size under clean data conditions, it becomes less reliable when the data contain outliers.

Increasing the time dimension ($T$) also enhances the performance of all estimators by providing more repeated observations over time. When holding $N = 200$, $\phi = 2$, and $\tau = 10\%$ fixed, increasing $T$ from 5 to 20 leads to a noticeable

TABLE 2 MSE and MAE values of non-robust and robust estimates when $N$ = 100.

| Estimators | $\phi = 0.5$ | | | | | | $\phi = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| $\tau = 0\%$ | | | | | | | | | | | | |
| NOFNB | 0.0808 | 0.3914 | 0.0338 | 0.2530 | 0.0155 | 0.1744 | 0.0381 | 0.2681 | 0.0160 | 0.1737 | 0.0072 | 0.1188 |
| FNBHM | 0.1134 | 0.4665 | 0.0517 | 0.3143 | 0.0227 | 0.2104 | 0.0472 | 0.2962 | 0.0205 | 0.1953 | 0.0095 | 0.1350 |
| FNBHR | 0.1537 | 0.5367 | 0.0850 | 0.4041 | 0.0504 | 0.3211 | 0.0707 | 0.3632 | 0.0385 | 0.2779 | 0.0254 | 0.2324 |
| FNBTK | 0.1839 | 0.5862 | 0.1062 | 0.4493 | 0.0611 | 0.3511 | 0.0746 | 0.3728 | 0.0396 | 0.2794 | 0.0255 | 0.2300 |
| $\tau = 5\%$ | | | | | | | | | | | | |
| NOFNB | 0.2543 | 0.6974 | 0.2110 | 0.6858 | 0.2270 | 0.7621 | 0.1669 | 0.6019 | 0.1725 | 0.6630 | 0.1742 | 0.6959 |
| FNBHM | 0.1495 | 0.5302 | 0.0925 | 0.4382 | 0.0813 | 0.4337 | 0.0523 | 0.3155 | 0.0332 | 0.2585 | 0.0213 | 0.2131 |
| FNBHR | 0.1238 | 0.4808 | 0.0566 | 0.3309 | 0.0349 | 0.2600 | 0.0512 | 0.3113 | 0.0281 | 0.2339 | 0.0161 | 0.1798 |
| FNBTK | 0.1543 | 0.5354 | 0.0741 | 0.3787 | 0.0504 | 0.3158 | 0.0548 | 0.3228 | 0.0308 | 0.2460 | 0.0177 | 0.1889 |
| $\tau = 10\%$ | | | | | | | | | | | | |
| NOFNB | 0.4991 | 1.0587 | 0.4620 | 1.0908 | 0.5239 | 1.2111 | 0.4134 | 1.0316 | 0.3967 | 1.0490 | 0.4816 | 1.1829 |
| FNBHM | 0.2150 | 0.6592 | 0.1562 | 0.5931 | 0.1454 | 0.6099 | 0.0970 | 0.4495 | 0.0728 | 0.4100 | 0.0691 | 0.4251 |
| FNBHR | 0.1820 | 0.5919 | 0.1128 | 0.4787 | 0.1004 | 0.4766 | 0.0998 | 0.4531 | 0.0743 | 0.4087 | 0.0789 | 0.4532 |
| FNBTK | 0.2144 | 0.6462 | 0.1344 | 0.5269 | 0.1211 | 0.5309 | 0.0905 | 0.4273 | 0.0668 | 0.3826 | 0.0670 | 0.4134 |
| $\tau = 15\%$ | | | | | | | | | | | | |
| NOFNB | 0.7167 | 1.3225 | 0.6996 | 1.3789 | 0.7808 | 1.4996 | 0.6425 | 1.3253 | 0.7232 | 1.4436 | 0.7494 | 1.4863 |
| FNBHM | 0.2774 | 0.7681 | 0.2397 | 0.7607 | 0.2238 | 0.7781 | 0.1745 | 0.6311 | 0.1494 | 0.6216 | 0.1490 | 0.6477 |
| FNBHR | 0.2675 | 0.7356 | 0.2143 | 0.6926 | 0.2082 | 0.7359 | 0.1842 | 0.6477 | 0.1755 | 0.6759 | 0.1789 | 0.7116 |
| FNBTK | 0.2773 | 0.7508 | 0.2239 | 0.7087 | 0.2080 | 0.7334 | 0.1451 | 0.5626 | 0.1275 | 0.5661 | 0.1299 | 0.6013 |
| $\tau = 20\%$ | | | | | | | | | | | | |
| NOFNB | 0.9523 | 1.5727 | 0.9892 | 1.6714 | 1.0328 | 1.7364 | 0.8998 | 1.5924 | 0.9594 | 1.6736 | 0.9959 | 1.7184 |
| FNBHM | 0.3823 | 0.9277 | 0.3201 | 0.9086 | 0.3176 | 0.9421 | 0.3110 | 0.8856 | 0.2948 | 0.9039 | 0.2863 | 0.9089 |
| FNBHR | 0.3993 | 0.9329 | 0.3571 | 0.9526 | 0.3594 | 0.9971 | 0.3239 | 0.9037 | 0.3243 | 0.9491 | 0.3269 | 0.9729 |
| FNBTK | 0.3820 | 0.9031 | 0.3142 | 0.8834 | 0.3098 | 0.9199 | 0.2430 | 0.7652 | 0.2294 | 0.7884 | 0.2238 | 0.7999 |

reduction in both MSE and MAE values. For instance, the MSE of the proposed estimator FNBHM decreases from ~0.0765 at $T = 5$ to 0.0597 at $T = 20$, reflecting a substantial improvement in estimation precision. This trend highlights the importance of longer time-series dimensions in reducing estimation error, even in the presence of moderate levels of outliers. We also find that, when comparing the performance across both under-dispersion ($\phi = 0.5$) and over-dispersion ($\phi = 2$) scenarios, the robust estimators generally maintain their superiority in the presence of outliers.

The MRE for robust estimators (FNBHR, FNBHM, and FNBTK) is illustrated in Figures 1, 2 clustered by percentages of outliers (from 5 to 20%) and time-series (from 5 to 20) for all units separately. The evaluation was performed separately for two cases; the first case expresses under-dispersion ($\phi = 0.5$), while the second case expresses over-dispersion ($\phi = 2$).

The results presented in Figures 1, 2 highlight the differences in estimators performance under varying dispersion levels. The efficiency of these estimators increases when increasing the size of the cross section ($N$), the time series ($T$), the percentage of the outliers ($\tau$), and the dispersion parameter ($\phi$), but the FNBTK estimator remains more efficient. Overall, the findings indicate that the MRE values of the FNBTK estimator are consistently higher than those of the FNBHR and FNBHM estimators. This suggests that the FNBTK estimator is more efficient than the other two robust estimators across all scenarios of $N$, $T$, $\tau$, and $\phi$.

## 4 Real-life panel data applications

The practical relevance of the proposed estimation methods is evaluated through real-life panel data applications from selected European countries, representing different types of count panel data frequently encountered in public health and economic contexts. To demonstrate the flexibility and robustness of these estimators under real-world conditions, two empirical case studies

TABLE 3 MSE and MAE values of non-robust and robust estimates when $N$ = 200.

| Estimators | $\phi = 0.5$ | | | | | | $\phi = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| $\tau = 0\%$ | | | | | | | | | | | | |
| NOFNB | 0.0373 | 0.2672 | 0.0171 | 0.1817 | 0.0077 | 0.1210 | 0.0174 | 0.1799 | 0.0076 | 0.1213 | 0.0035 | 0.0824 |
| FNBHM | 0.0532 | 0.3196 | 0.0252 | 0.2217 | 0.0127 | 0.1565 | 0.0221 | 0.2005 | 0.0100 | 0.1390 | 0.0049 | 0.0964 |
| FNBHR | 0.0897 | 0.4190 | 0.0567 | 0.3429 | 0.0406 | 0.3025 | 0.0393 | 0.2739 | 0.0245 | 0.2266 | 0.0201 | 0.2188 |
| FNBTK | 0.1057 | 0.4512 | 0.0676 | 0.3717 | 0.0498 | 0.3336 | 0.0406 | 0.2773 | 0.0247 | 0.2256 | 0.0196 | 0.2129 |
| $\tau = 5\%$ | | | | | | | | | | | | |
| NOFNB | 0.1948 | 0.6545 | 0.1931 | 0.6942 | 0.2021 | 0.7444 | 0.1411 | 0.5880 | 0.1731 | 0.6880 | 0.1697 | 0.6983 |
| FNBHM | 0.0951 | 0.4437 | 0.0720 | 0.3992 | 0.0659 | 0.4071 | 0.0309 | 0.2454 | 0.0235 | 0.2203 | 0.0181 | 0.2062 |
| FNBHR | 0.0598 | 0.3394 | 0.0312 | 0.2423 | 0.0196 | 0.1951 | 0.0257 | 0.2218 | 0.0181 | 0.1878 | 0.0120 | 0.1591 |
| FNBTK | 0.0771 | 0.3880 | 0.0428 | 0.2842 | 0.0309 | 0.2505 | 0.0277 | 0.2303 | 0.0200 | 0.1981 | 0.0135 | 0.1702 |
| $\tau = 10\%$ | | | | | | | | | | | | |
| NOFNB | 0.4379 | 1.0524 | 0.4828 | 1.1581 | 0.4923 | 1.1928 | 0.3734 | 1.0152 | 0.4618 | 1.1578 | 0.4512 | 1.1541 |
| FNBHM | 0.1576 | 0.5919 | 0.1435 | 0.6026 | 0.1323 | 0.6033 | 0.0765 | 0.4204 | 0.0677 | 0.4189 | 0.0597 | 0.4066 |
| FNBHR | 0.1104 | 0.4711 | 0.0916 | 0.4523 | 0.0827 | 0.4541 | 0.0736 | 0.4059 | 0.0742 | 0.4372 | 0.0663 | 0.4278 |
| FNBTK | 0.1342 | 0.5241 | 0.1156 | 0.5136 | 0.1018 | 0.5098 | 0.0652 | 0.3792 | 0.0632 | 0.3983 | 0.0566 | 0.3926 |
| $\tau = 15\%$ | | | | | | | | | | | | |
| NOFNB | 0.6628 | 1.3423 | 0.7112 | 1.4265 | 0.7645 | 1.4985 | 0.6101 | 1.3201 | 0.6849 | 1.4196 | 0.7287 | 1.4716 |
| FNBHM | 0.2354 | 0.7507 | 0.2119 | 0.7518 | 0.2129 | 0.7778 | 0.1607 | 0.6441 | 0.1507 | 0.6500 | 0.1371 | 0.6291 |
| FNBHR | 0.2037 | 0.6731 | 0.1892 | 0.6928 | 0.1966 | 0.7389 | 0.1642 | 0.6480 | 0.1673 | 0.6847 | 0.1671 | 0.6959 |
| FNBTK | 0.2182 | 0.6976 | 0.1891 | 0.6906 | 0.1945 | 0.7343 | 0.1262 | 0.5590 | 0.1241 | 0.5846 | 0.1220 | 0.5914 |
| $\tau = 20\%$ | | | | | | | | | | | | |
| NOFNB | 0.8821 | 1.5684 | 0.9585 | 1.6683 | 1.0116 | 1.0116 | 0.8371 | 1.5583 | 0.9383 | 1.6663 | 0.9736 | 1.7035 |
| FNBHM | 0.3141 | 0.8914 | 0.3119 | 0.9275 | 0.3040 | 0.3040 | 0.2854 | 0.8861 | 0.2833 | 0.9020 | 0.2711 | 0.8928 |
| FNBHR | 0.3209 | 0.8846 | 0.3360 | 0.9546 | 0.3481 | 0.3481 | 0.2915 | 0.8941 | 0.3105 | 0.9451 | 0.3118 | 0.9581 |
| FNBTK | 0.3030 | 0.8544 | 0.2977 | 0.8935 | 0.2948 | 0.2948 | 0.2155 | 0.7591 | 0.2209 | 0.7915 | 0.2143 | 0.7921 |

are presented. The first application utilizes COVID-19 data, while the second focuses on patent data, providing insights into the dynamics of innovation and economic development over time. Both datasets exhibit the panel structure necessary to evaluate the performance of classical and robust estimators under varying degrees of heterogeneity, over-dispersion, and the presence of outliers. Through these applications, the practical value and adaptability of the proposed methods are highlighted across real-world scenarios.

## 4.1 COVID-19 application

In recent years, the global emergence of COVID-19 has led to profound disturbances in healthcare systems, economic structures, and social dynamics worldwide. The availability of COVID-19-related data has played a pivotal role in facilitating research focused on understanding the virus's spread and evaluating the effectiveness of public health actions. Such data have proven essential for researchers, policymakers, and healthcare institutions striving to mitigate the pandemic's impact and address its multifaceted challenges.

Comprehensive and reliable COVID-19 datasets have provided a solid foundation for conducting empirical research. These datasets have not only enhanced our understanding of pandemic dynamics but have also spurred advancements in statistical modeling, particularly in the fields of count panel data analysis, panel data techniques, and robust estimation approaches. As a result, COVID-19 data have become instrumental in applying and validating modern statistical methods, promoting innovation in methodology, encouraging interdisciplinary collaboration, and contributing to the design of evidence-based policies. In this context, the use of such data highlights the practical relevance and empirical strength of the proposed estimation techniques,

TABLE 4 MSE and MAE values of non-robust and robust estimates when $N$ = 400.

| Estimators | $\phi = 0.5$ | | | | | | $\phi = 2$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | | $T$ = 5 | | $T$ = 10 | | $T$ = 20 | |
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| $\tau = 0\%$ | | | | | | | | | | | | |
| NOFNB | 0.0182 | 0.1869 | 0.0081 | 0.1239 | 0.0038 | 0.0850 | 0.0084 | 0.1269 | 0.0038 | 0.0850 | 0.0017 | 0.0584 |
| FNBHM | 0.0263 | 0.2269 | 0.0127 | 0.1536 | 0.0065 | 0.1125 | 0.0108 | 0.1425 | 0.0052 | 0.0996 | 0.0029 | 0.0739 |
| FNBHR | 0.0620 | 0.3589 | 0.0415 | 0.3090 | 0.0373 | 0.3091 | 0.0267 | 0.2362 | 0.0205 | 0.2195 | 0.0187 | 0.2209 |
| FNBTK | 0.0725 | 0.3849 | 0.0513 | 0.3422 | 0.0441 | 0.3343 | 0.0265 | 0.2323 | 0.0203 | 0.2165 | 0.0178 | 0.2127 |
| $\tau = 5\%$ | | | | | | | | | | | | |
| NOFNB | 0.1857 | 0.6746 | 0.1878 | 0.7162 | 0.1942 | 0.7458 | 0.1458 | 0.6292 | 0.1565 | 0.6693 | 0.1622 | 0.6900 |
| FNBHM | 0.0732 | 0.3965 | 0.0621 | 0.3943 | 0.0599 | 0.4021 | 0.0216 | 0.2108 | 0.0173 | 0.2008 | 0.0152 | 0.1984 |
| FNBHR | 0.0328 | 0.2486 | 0.0176 | 0.1861 | 0.0128 | 0.1609 | 0.0154 | 0.1724 | 0.0110 | 0.1516 | 0.0088 | 0.1412 |
| FNBTK | 0.0452 | 0.2918 | 0.0272 | 0.2342 | 0.0228 | 0.2215 | 0.0169 | 0.1804 | 0.0123 | 0.1614 | 0.0102 | 0.1537 |
| $\tau = 10\%$ | | | | | | | | | | | | |
| NOFNB | 0.4411 | 1.1003 | 0.4481 | 1.1373 | 0.4740 | 1.1819 | 0.3929 | 1.0641 | 0.4348 | 1.1310 | 0.4414 | 1.1461 |
| FNBHM | 0.1361 | 0.5794 | 0.1294 | 0.5951 | 0.1272 | 0.6054 | 0.0633 | 0.4007 | 0.0606 | 0.4083 | 0.0577 | 0.4074 |
| FNBHR | 0.0844 | 0.4254 | 0.0744 | 0.4268 | 0.0727 | 0.4459 | 0.0633 | 0.3960 | 0.0646 | 0.4198 | 0.0630 | 0.4252 |
| FNBTK | 0.1057 | 0.4817 | 0.0942 | 0.4863 | 0.0926 | 0.5066 | 0.0545 | 0.3627 | 0.0543 | 0.3820 | 0.0535 | 0.3904 |
| $\tau = 15\%$ | | | | | | | | | | | | |
| NOFNB | 0.6922 | 1.4044 | 0.7440 | 1.4777 | 0.7476 | 1.4897 | 0.6447 | 1.3744 | 0.6519 | 1.3908 | 0.7135 | 1.4595 |
| FNBHM | 0.2111 | 0.7447 | 0.2110 | 0.7728 | 0.2007 | 0.7657 | 0.1462 | 0.6374 | 0.1390 | 0.6333 | 0.1362 | 0.6334 |
| FNBHR | 0.1853 | 0.6771 | 0.1877 | 0.7190 | 0.1816 | 0.7231 | 0.1576 | 0.6607 | 0.1529 | 0.6642 | 0.1616 | 0.6902 |
| FNBTK | 0.1905 | 0.6860 | 0.1924 | 0.7274 | 0.1801 | 0.7200 | 0.1189 | 0.5673 | 0.1144 | 0.5717 | 0.1175 | 0.5872 |
| $\tau = 20\%$ | | | | | | | | | | | | |
| NOFNB | 0.9293 | 1.6415 | 0.9404 | 1.6666 | 0.9907 | 1.7178 | 0.8854 | 1.6168 | 0.9786 | 1.7077 | 0.9630 | 1.6969 |
| FNBHM | 0.3021 | 0.9098 | 0.2872 | 0.9089 | 0.2938 | 0.9297 | 0.2797 | 0.8961 | 0.2810 | 0.9086 | 0.2732 | 0.9007 |
| FNBHR | 0.3228 | 0.9329 | 0.3102 | 0.9408 | 0.3332 | 0.9888 | 0.2929 | 0.9165 | 0.3134 | 0.9601 | 0.3072 | 0.9553 |
| FNBTK | 0.2907 | 0.8799 | 0.2744 | 0.8819 | 0.2836 | 0.9107 | 0.2124 | 0.7751 | 0.2158 | 0.7941 | 0.2117 | 0.7918 |

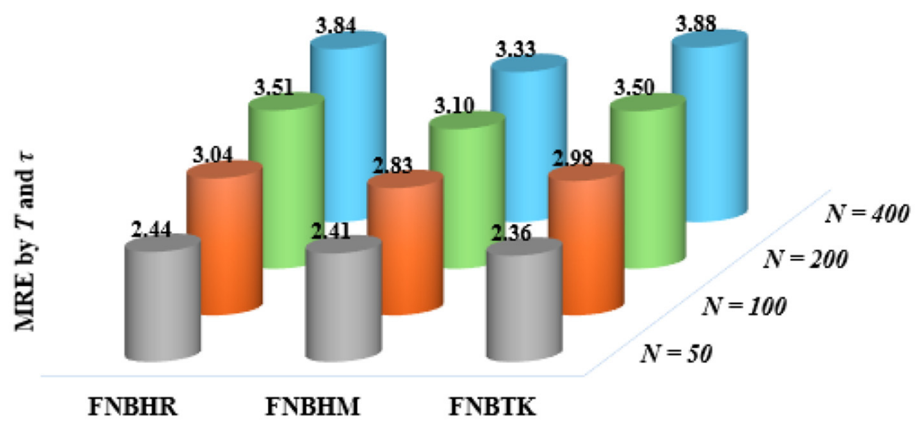


FIGURE 1
MRE of the robust estimates when $\phi = 0.5$.

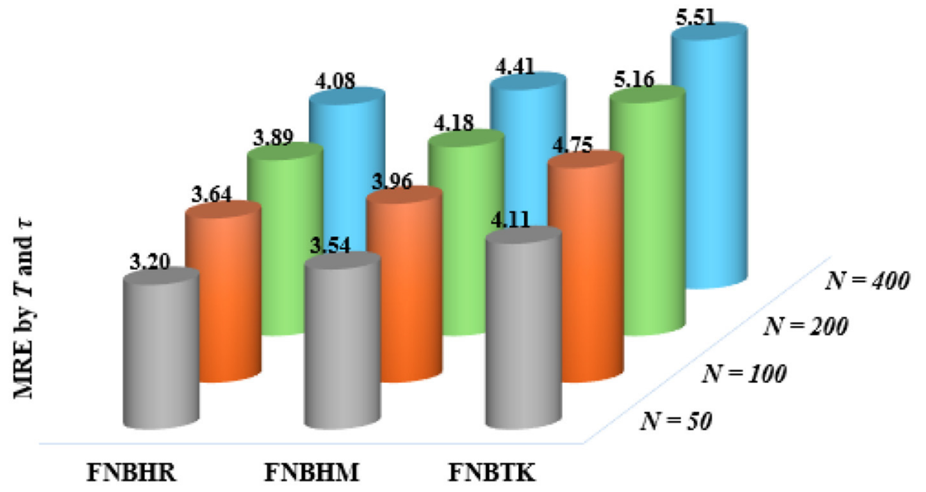


FIGURE 2
MRE of the robust estimates when $\phi = 2$.

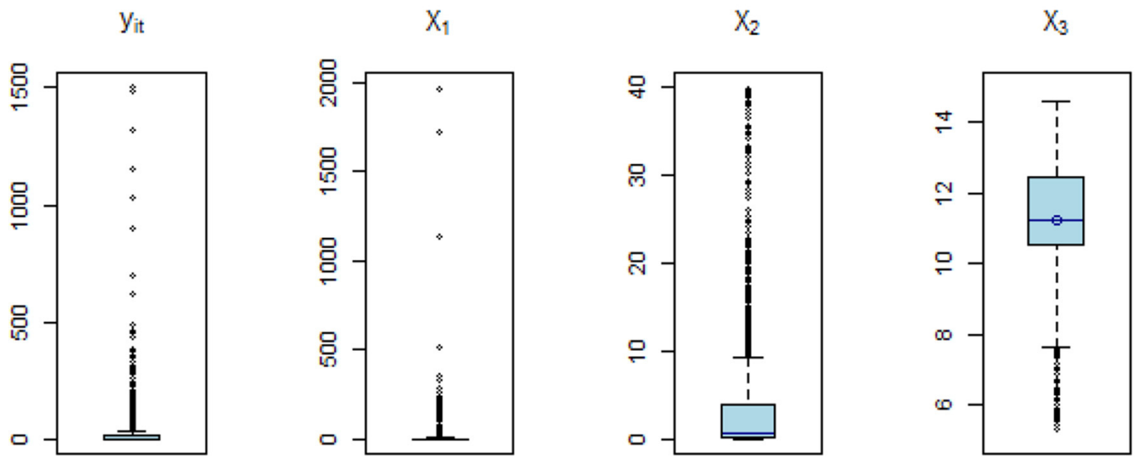


FIGURE 3
Box plots of COVID-19 variables containing outliers.

especially in addressing real-world problems associated with public health crises.

This application employs a daily panel dataset on COVID-19, covering ten European countries during the period from June 23, 2021, to January 21, 2022. The data were obtained from the World Health Organization website. The analysis includes five variables, with the response variable being the number of new COVID-19-related deaths ($y_{it}$). The explanatory variables are the number of confirmed cases per 1,000 individuals ($X_1$), the number of patients in intensive care units per 100 individuals ($X_2$), the logarithm of new COVID-19 tests ($X_3$), and the logarithm of the number of individuals who received COVID-19 vaccinations ($X_4$).

Figure 3 presents boxplots of the study variables that exhibit outliers across the selected countries. A visual examination of the boxplots reveals the presence of outliers in the $y_{it}$ as well as in the explanatory variables: $X_1$, $X_2$, and $X_3$. Furthermore, an over-dispersion test was conducted, and the results indicated a statistically significant presence of over-dispersion in the data, with a $p$-value $< 0.001$.

The algorithm described in Section 2 is applied to obtain the estimators' results, aiming to derive more accurate estimates and enhance the robustness and reliability of the analysis. Table 5 presents the coefficient estimates from the NB panel model using both non-robust and robust methods. The results in Table 5 summarize the estimated coefficients for the NB panel model applied to the COVID-19 dataset. The NOFNB estimator, which does not account for outliers, produces significantly different coefficient values compared to the robust estimators.

To evaluate the performance of these estimators, we refer to the goodness-of-fit measures reported in Table 6. These measures clearly show that the robust estimators yield lower values for AIC and BIC, demonstrating their superior performance. Among them, the FNBTK estimator achieves the lowest values across all criteria, confirming its effectiveness in handling outliers in count panel data.

TABLE 5 Coefficient estimates for various estimators using the COVID-19 dataset.

| Variables | NOFNB | | FNBHM | | FNBHR | | FNBTK | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE |
| Intercept | 2.611* | 0.043 | 1.011* | 0.043 | 1.276* | 0.049 | 0.759* | 0.045 |
| $X_1$ | 0.012* | 0.001 | 0.225* | 0.011 | 0.106* | 0.009 | 0.249* | 0.013 |
| $X_2$ | 0.172* | 0.014 | -0.497* | 0.046 | 0.001 | 0.041 | −0.597* | 0.047 |
| $X_3$ | −0.860* | 0.075 | -0.486* | 0.066 | -0.591* | 0.077 | −0.538* | 0.067 |
| $X_4$ | −0.329* | 0.040 | -0.259* | 0.033 | −0.289* | 0.038 | −0.250* | 0.033 |

The symbols (*) indicate that the significance level at < 1%.

TABLE 6 Goodness-of-fit measures for the estimators in COVID-19 dataset.

| Estimators | Log-likelihood | AIC | BIC |
|---|---|---|---|
| NOFNB | −6,794 | 13,599 | 13,633 |
| FNBHM | −2,882 | 5,777 | 5,811 |
| FNBHR | −2,960 | 5,933 | 5,967 |
| FNBTK | −2,346 | 4,703 | 4,737 |

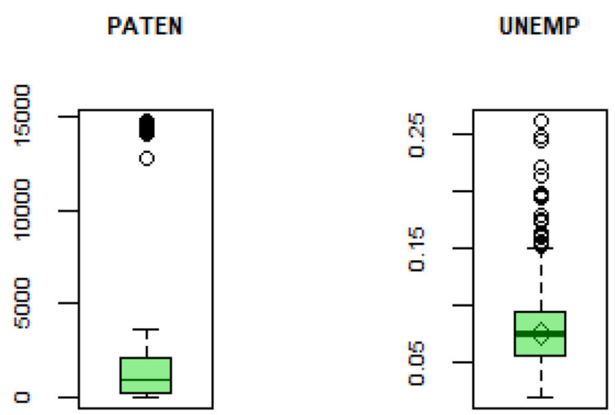


FIGURE 4
Box plots of patent variables containing outliers.

## 4.2 Patents application

Another empirical application is conducted to evaluate the performance of the proposed estimators using patents data. In the global economy, patents stand as cornerstones of innovation, economic growth, and competitive advantage, particularly in high-income European countries. These countries have long recognized the real value of intellectual property rights in creating an environment that supports and promotes research, development, and marketing the new ideas. Patents, by providing exclusive rights to inventors for a defined period, encourage the development and dissemination of innovative technologies, thus propelling advancements in various sectors including healthcare, information technology, and green energy. Moreover, in the context of the European single market, patents play a pivotal role in harmonizing standards, facilitating cross-border trade, and enhancing the region's attractiveness to investors and entrepreneurs from around the globe. Consequently, understanding the importance of patent protection in these countries is a crucial element in comprehending their economic resilience and their ability to find pioneering solutions to global challenges. In this context, we will employ the patents dataset to evaluate various estimators.

In our application, we are interested in estimating patent applications. As defined by the World Intellectual Property Organization, a patent application is associated with either a new method of operation or a novel technical problem solution. We based our sample selection on the patent data for high-income European countries available on the World Bank website. This sample includes 17 countries classified as high-income, covering the years from 2005 until 2020. In this analysis, we focus on the patents (*PATEN*) as the dependent variable. The independent variables are the number of researchers in research and development (*NUMRD*) per 1000 people, the logarithm of the research and development expenditures (*RDEXD*), the logarithm of the GDP per capita (*GDPPC*), the logarithm of the exports of technological goods (*ICTEXP*), and the rate of unemployment (*UNEMP*).

Figure 4 displays box plots for the variables that exhibit outliers across the countries under study. Specifically, the box plots indicate that the response variable and UNEMP contain outlier values, while the remaining variables do not show apparent outliers. Also, we conducted an over-dispersion test and the results indicated a significant presence of over-dispersion in the data < 0.001 significance level. Table 7 shows the findings of the NB panel model of non-robust and robust estimates.

After assessing the AIC and BIC criteria, as illustrated in Table 8, to a conclusion that robust estimates for the NB panel model outperformed non-robust estimates in terms of effectiveness, which indicates that the robust estimators FNBHM, FNBHR, and FNBTK outperform when compared to the NOFNB non-robust estimator. The criterion for determining the best estimator relies on selecting the one with the lowest value for either AIC or BIC. This finding emphasizes the importance of selecting robust estimators for enhanced accuracy and reliability in evaluating the results. In addition, the FNBTK estimator demonstrates better performance compared to the other estimators. The findings clearly demonstrate that robust estimators outperform classical estimator when outliers are present.

TABLE 7 Coefficient estimates for various estimators using the patent dataset.

| Variables | NOFNB | | FNBHM | | FNBHR | | FNBTK | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE | Estimate | SE |
| Intercept | −12.589* | 1.730 | −12.089* | 1.824 | −11.875* | 1.859 | −11.650* | 1.930 |
| NUMRD | −1.108* | 0.170 | −1.155* | 0.194 | −1.207* | 0.207 | −1.236* | 0.228 |
| RDEXD | 0.908* | 0.064 | 0.901* | 0.068 | 0.904* | 0.069 | 0.901* | 0.072 |
| GDPPC | −0.381* | 0.093 | −0.382* | 0.094 | −0.382* | 0.095 | −0.384* | 0.097 |
| ICTEXP | 0.011 | 0.055 | −0.002 | 0.058 | −0.013 | 0.059 | −0.019 | 0.060 |
| UNEMP | 4.058 | 3.737 | 4.236 | 3.735 | 4.083 | 3.729 | 4.475 | 3.799 |

The symbols (*) indicate that the significance level at $< 1\%$.

TABLE 8 Goodness-of-fit measures for the estimators in patent dataset.

| Estimators | Log-likelihood | AIC | BIC |
|---|---|---|---|
| NOFNB | −415 | 843 | 857 |
| FNBHM | −380 | 774 | 788 |
| FNBHR | −358 | 729 | 743 |
| FNBTK | −333 | 680 | 694 |

## 5 Conclusion

In this research study, we have presented proposed robust estimators for the FENB panel model. To evaluate the performance of these estimators, we conducted the simulation study and applied them to real panel datasets. By combining Monte Carlo simulation and the applications of real data, we can conduct a comprehensive evaluation of the proposed estimators, ensuring their effectiveness and reliability. This dual approach allows us to assess the estimators' performance across different situations, providing a comprehensive understanding of their robustness and practical utility under diverse conditions.

The simulation results indicate that in the absence of outliers, the NOFNB estimator outperforms the robust estimators, but, when outliers are present, the proposed robust estimators (FNBHM, FNBHR, and FNBTK) exhibit higher effectiveness compared to the non-robust estimator. Furthermore, the results of the empirical studies in European countries indicate that the proposed robust estimators are more efficient than other estimators when outliers are present in the datasets. In addition, the FNBTK robust estimator demonstrates greater efficiency compared to FNBHM and FNBHR robust estimators.

The results of this study are expected to have practical implications for researchers and policymakers in the field of scientific research and development. In addition, it contributes to the advancement of statistical methodology for analyzing count panel data in the presence of outliers and provides a better understanding of the factors influencing count panel data in different fields by emphasizing the importance of the NB panel model. In future work, we plan to propose new robust estimators for the FENB model or for one of the other count panel data models.

## Data availability statement

Publicly available datasets were analyzed in this study. The first data can be found at: https://www.who.int/data. The second data can be found at: https://databank.worldbank.org/. The files for both datasets are found in the Supplementary material.

## Author contributions

MA: Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. EE: Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – review & editing. EA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. AE-M: Data curation, Funding acquisition, Investigation, Project administration, Resources, Software, Supervision, Visualization, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2025.1638596/full#supplementary-material

## References

1. Ahmed EG, Abonazel MR, Al-Ghamdi MN, Elshamy HM, Khattab IG. Proposed robust estimators for the Poisson panel regression model: application to COVID-19 deaths in Europe. *Commun Math Biol Neurosci.* (2024) 2024:121. doi: 10.28919/cmbn/8795

2. Abonazel MR, Youssef AH, Ahmed EG. On robust M, S and MM estimations for the Poisson fixed effects panel model with outliers: simulation and applications. *J Stat Comput Simul.* (2025) 95:1190–207. doi: 10.1080/00949655.2024.2449102

3. Youssef AH, Abonazel MR, Ahmed EG. Robust M estimation for Poisson panel data model with fixed effects: method, algorithm, simulation, and application. *Stat Optim Inf Comput.* (2024) 12:1292–305. doi: 10.19139/soic-2310-5070-1996

4. Baltagi BH. *Econometric Analysis of Panel Data*. Hoboken, NJ: John Wiley and Sons Ltd (2005).

5. Abonazel MR, Ahmed EG, El-Baz ST, Ebrahim EEM. How economic policies drive climate change: a comparative analysis of groups of middle- and high-income countries. *Int J Energy Econ Policy.* (2025) 15:447–55. doi: 10.32479/ijeep.17853

6. Winkelmann R. *Econometric Analysis of Count Data*. Berlin: Springer-Verlag. (2008).

7. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press (2013). doi: 10.1017/CBO9781139013567

8. Yaacob WFW, Lazim MA, Wah YB. A practical approach in modelling count data. In: *Proceedings of the Regional Conference on Statistical Sciences*. Shah Alam: Malaysia Institute of Statistics, Faculty of Computer and Mathematical Sciences, University Technology MARA (UiTM) (2010). p. 176–183.

9. Hana N. Negative binomial regression model for road crash severity prediction. *Mod Appl Sci.* (2018) 12:38–46. doi: 10.5539/mas.v12n4p38

10. Al-Taweel Y, Algamal Z. Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model. *Period Eng Nat Sci.* (2020) 8:248–55. doi: 10.21533/pen.v8i1.1107

11. Alobaidi NN, Shamany RE, Algamal ZY. A new ridge estimator for the negative binomial regression model. *Thailand Stat.* (2021) 19:115–24.

12. Rashad NK, Hammood NM, Algamal ZY. Generalized ridge estimator in negative binomial regression model. *J Phys Conf Ser.* (2021) 1897:012019. doi: 10.1088/1742-6596/1897/1/012019

13. Niang MA, Diop A, Seck CT, Faye I, Ndao A, Sy IA, et al. negative binomial regression analysis for COVID-19 death cases in Senegal. *Int J Stat Appl Math.* (2022) 7:6–12. doi: 10.22271/maths.2022.v7.i3a.817

14. Liu R, Heo I, Liu H, Shi D, Jiang Z. Applying negative binomial distribution in diagnostic classification models for analyzing count data. *Appl Psychol Meas.* (2023) 47:64–75. doi: 10.1177/01466216221124604

15. Suryadi F, Jonathan S, Jonatan K, Ohyver M. Handling overdispersion in Poisson regression using negative binomial regression for poverty case in West Java. *Procedia Comput Sci.* (2023) 216:517–23. doi: 10.1016/j.procs.2022.12.164

16. Youssef AH, Abonazel MR, Ahmed EG. Estimating the number of patents in the world using count panel data models. *Asian J Probab Stat.* (2020) 6:24–33. doi: 10.9734/ajpas/2020/v6i430167

17. Kumara SSP, Chin HC. Study of fatal traffic accidents in Asia Pacific countries. *Transp Res Rec.* (2004) 1897:43–7. doi: 10.3141/1897-06

18. Guimaraes P. The fixed effects negative binomial model revisited. *Econ Lett.* (2008) 99:63–6. doi: 10.1016/j.econlet.2007.05.030

19. Hausman JA, Hall BH, Griliches Z. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica.* (1984) 52:909–38. doi: 10.2307/1911191

20. Youssef AH, Abonazel MR, Ahmed EG. The performance of count panel data estimators: a simulation study and application to patents in Arab countries. *J Math Comput Sci.* (2021) 11:8173–96. doi: 10.28919/jmcs/5852

21. Huber P. Robust estimation of a location parameter. *Ann Math Stat.* (1964) 35:73–101. doi: 10.1214/aoms/1177703732

22. Huber PJ. *Robust Statistics*. Hobokn, NJ: John Wiley & Sons (1981). doi: 10.1002/0471725250

23. Rousseeuw PJ, Leroy AM. *Robust Regression and Outlier Detection*. Hobokn, NJ: John Wiley & Sons (1987). doi: 10.1002/0471725382

24. Garay AM, Hashimoto EM, Ortega EMM, Lachos VH. On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Comput Stat Data Anal.* (2011) 55:1304–18. doi: 10.1016/j.csda.2010.09.019

25. Toka O, Cetin M. The comparing of S-estimator and M-estimators in linear regression. *Gazi Univ J Sci.* (2011) 24:747–52.

26. Susanti Y, Pratiwi H, Sulistijowati S, Liana T. M estimation, S estimation, and MM estimation in robust regression. *Int J Pure Appl Math.* (2014) 91:349–60. doi: 10.12732/ijpam.v91i3.7

27. Maronna RA, Martin RD, Yohai VJ. *Robust Statistics: Theory and Methods*. Hoboken, NJ: John Wiley & Sons (2019). doi: 10.1002/9781119214656

28. Tüzen F, ErbaŞ, OlmuŞ H. A simulation study for count data models under varying degrees of outliers and zeros. *Commun Stat-Simul Comput.* (2020) 49:1078–1088. doi: 10.1080/03610918.2018.1498886

29. Youssef AH, Abonazel MR, Kamel AR. Efficiency comparisons of robust and non-robust estimators for seemingly unrelated regressions model. *WSEAS Trans Math.* (2022) 21:218–44. doi: 10.37394/23206.2022.21.28

30. Ke S, Phillips PC, Su L. Robust inference of panel data models with interactive fixed effects under long memory: a frequency domain approach. *J Econom.* (2024) 241:105761. doi: 10.1016/j.jeconom.2024.105761

31. Beyaztas BH, Bandyopadhyay S. Data driven robust estimation methods for fixed effects panel data models. *J Stat Comput Simul.* (2022) 92:1401–25. doi: 10.1080/00949655.2021.1996576

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

33. Víšek JÁ. Estimating the model with fixed and random effects by a robust method. *Methodol Comput Appl Probab.* (2015) 17:999–1014. doi: 10.1007/s11009-014-9432-5

34. Bramati MC, Croux C. Robust estimators for the fixed effects panel data model. *Econometr J.* (2007) 10:521–40. doi: 10.1111/j.1368-423X.2007.00220.x

35. Hampel F, Ronchetti E, Rousseeuw P, Stahel W. *Robust Statistics: The Approach Based on Influence Functions*. Hoboken, NJ: John Wiley & Sons (1986).

36. Venables WN, Ripley BD. *Modern Applied Statistics with S*, 4th Edn. Cham: Springer (2002). doi: 10.1007/978-0-387-21706-2

37. Baltagi BH. *The Oxford Handbook of Panel Data. Oxford Handbooks*. Oxford: Oxford University Press (2015). doi: 10.1093/oxfordhb/9780199940042.001.0001