

Robust M Estimation for Poisson Panel Data Model with Fixed Effects: Method, Algorithm, Simulation, and Application

Ahmed H. Youssef¹, Mohamed R. Abonazel¹, Elsayed G. Ahmed^{1,2,*}

¹*Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt*

²*Department of Statistics and Insurance, Faculty of Commerce, Arish University, North Sinai, Egypt*

Abstract The fixed effects Poisson (FEP) model is crucial for count data involving periods and cross-sectional units. The maximum likelihood (ML) estimation method for the FEP model performs well without outliers, but its performance degrades in their presence. Therefore, this paper introduces robust estimators for the FEP model. These estimators provide stable and reliable results even when outliers are present. A Monte Carlo simulation study and an empirical application were conducted to evaluate the performance of the non-robust fixed Poisson maximum likelihood (FPML) estimator and the robust estimators: fixed Poisson Huber (FPHR), fixed Poisson Hampel (FPHM), and fixed Poisson Tukey (FPTK). The findings from the simulation and application indicate that robust estimators outperform the FPML estimator in the presence of outliers in count panel data. Furthermore, the FPTK estimator is more efficient than the other robust estimators.

Keywords Count Data, Outliers, Fixed Effects Poisson Model, Maximum Likelihood, M Estimation, Monte Carlo Simulation, and Empirical Application.

AMS 2010 subject classifications: 62J12, 62J20, 62F35

DOI: 10.19139/soic-2310-5070-1996

1. Introduction

Count data refers to observations that take nonnegative integer values, these values are generated from counting because of the occurrence of an event, for example, the number of accidents that occur on the roads, the number of patents granted to countries and companies, and the number of deaths due to a particular disease. Count data are widely used in health and economic fields, so researchers are interested in models that deal with this type of data. In these models, the count data are treated as a response variable [34].

In the econometric literature, the Poisson regression model has been widely used to analyze count data, for example, in the field of healthcare and medicine, [4] used the Poisson regression model of count to model daily death cases of COVID-19 in Nigeria. There are many studies interested in the application of the Poisson regression model in other fields, see, e.g. [21, 32, 35, 30, 29, 8].

In recent years, panel data analysis has become one of the most exciting fields in the econometrics literature, where panel data aim to study the data of the time dimension and the sectoral dimension to achieve the maximum benefit from the data. According to [9], the panel dataset refers to the combining of observations on cross-section and time-series, where the cross-section is observed over several periods.

Panel data regression models have become widely used among applied researchers due to their multiple advantages compared to cross-section or time-series data models. Therefore, we will discuss one of the most

*Correspondence to: Elsayed Gamal Ahmed (Email: Elsayed.gamal@comm.aru.edu.eg). Department of Statistics and Insurance, Faculty of Commerce, Arish University, North Sinai, Egypt.

important models used in panel data regression modeling, which is the fixed effects model. Many studies used the fixed effects model; see, e.g. [1, 27, 37, 36, 38, 16, 5].

Despite the various models for count data regression, more progress has been made in improving the count data models by using panel data regression models. In count panel data models, the dependent variable does not follow the normal distribution, where the dependent variable takes nonnegative integer values. [18] introduced a variety of econometric models to deal with count data in panel data models, where the Poisson model with fixed effects is used by using the ML estimation method. For more studies and examples on count panel data in the fields of econometrics, political, biological, and health sciences, see, for example, [39, 7, 28, 17, 23, 26, 22, 14, 15].

The existing studies for the robust estimators of the Poisson regression model are limited, where some studies deal with outliers in the generalized linear models; see [20, 25, 3, 24, 2]. These studies introduced the Poisson model for the data that contain outliers. The Monte Carlo simulation was performed to compare the performance of the non-robust ML estimator with some robust estimators when the data suffer from the presence of outliers. [11] developed a robust generalized quasi-likelihood estimation method to estimate the parameters in longitudinal models for binary and count data with outliers. The simulation results show that the robust generalized quasi-likelihood produces unbiased and consistent estimates compared to the classical generalized quasi-likelihood estimation method.

Although panel data regression models are important, panel data often suffer from data outliers. These outliers have adverse effects on classical estimation methods, where the studies of robust estimation methods for fixed-effects panel data regression model are a few. For studies on robust methods for the fixed effects model, see [12, 6, 33, 13].

In the econometric literature, there is no specific robust method for estimating parameters in FEP. Therefore, in this paper, we introduce robust estimation methods for the FEP model in case the count panel data contains outliers based on the weight function for Huber, Hampel, and Tukey Bisquare.

2. Fixed Effects Poisson and ML Estimator

The Poisson panel model is a discrete probability distribution of the count of randomly occurring events for individuals i in the time t . The probability mass function of the FEP model is given by:

$$f(y_{it}|X_{k,it}, \alpha_i, \beta) = \frac{\exp(-\mu_{it})(\mu_{it})^{y_{it}}}{y_{it}!}, \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T \quad (1)$$

where the dependent variable, or y_{it} values are nonnegative integers, i.e., the model allows for the possibility of counts where $y_{it} \geq 0$. Based on y_{it} , we can write the model in terms of the mean of the response as:

$$y_{it} = E(y_{it}|X_{k,it}, \alpha_i, \beta) + u_{it} \quad (2)$$

where y_{it} is the dependent variable for individuals i at time t , $X_{k,it}$ is the it^{th} observation on k explanatory variables, α_i represent constant term for cross-sections and differing from unit to unit and fixed over time. The intercept α_i includes the unobserved effect for special variables to the i^{th} individual over time, β is the vector of the regression coefficients, and u_{it} is the error term of the model.

We can use the ML estimation method to estimate the regression coefficients of the model (1). The joint probability function for the i^{th} observation for model (1) is given by:

$$f(y_{it}|\alpha_i, \beta, X_{k,it}) = \prod_{t=1}^T \left\{ \frac{\exp(-\mu_{it})}{y_{it}!} \times (\mu_{it})^{y_{it}} \right\} \quad (3)$$

where $\mu_{it} = \exp(\delta_i + X'_{k,it}\beta)$, taking the logarithm of joint probability function in (3) and summing over all individuals, the log-likelihood function of individual i is:

$$L(\alpha_i, \beta) = -\alpha_i T \bar{\lambda}_i + \delta_i T \bar{y}_i + T \bar{y}_i \ln \lambda_{it} - \sum_{t=1}^T \ln y_{it}! \quad (4)$$

where $\delta_i = \ln \alpha_i$, $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$, and $\bar{\lambda}_i = \frac{1}{T} \sum_{t=1}^T \lambda_{it}$. Differentiating with respect to α_i and setting to zero, we get the ML estimator for α_i :

$$\hat{\alpha}_i = T \bar{y}_i \left[\sum_{t=1}^T \exp(X'_{k,it} \beta) \right]^{-1} \quad (5)$$

With the first derivative with respect to β and setting to zero for (4), taking into consideration all N observations, we get:

$$\sum_{i=1}^N \sum_{t=1}^T X'_{(k,it)} (y_{it} - \mu_{it}) = 0 \quad (6)$$

The ML estimates for β have no closed-form solution, so numerical search procedures are used to find the ML estimates for β in (6).

3. Proposed Robust Estimators

Robust regression for the Poisson panel model is an important tool for analyzing count panel data and providing good results in the presence of outliers. The ML estimates for (6) can be affected when the count panel data contains outliers. Therefore, we use the M estimation method to obtain stable estimations for the coefficients in the FEP model.

[19] generalized the median to a larger class of estimators, called M estimators (or ML-type estimators). The M estimation method is based on minimizing the residuals (disturbances) function, the residuals or disturbances corresponding to the observation it^{th} in the FEP model are:

$$R_{it}(\alpha_i, \beta) = y_{it} - \exp(\delta_i + X'_{(k,it)} \beta).$$

We can write the M estimator for the FEP model by minimizing the objective function ρ over all β as follows:

$$\hat{\beta}_M = \min \sum_{i=1}^N \sum_{t=1}^T \rho \left(\frac{\xi_{it}}{\hat{\sigma}} \right) \quad (7)$$

where $\rho(\cdot)$ is a continuous and symmetric objective function that satisfies certain properties. Often $\rho(\cdot)$ can be formed by using a generalized linear combination of the residuals. $\xi_{it} = \frac{R_{it}(\alpha_i, \beta)}{\sqrt{\hat{\alpha}_i \lambda_{it}}}$, $\lambda_{it} = \exp(X'_{(k,it)} \beta)$, and $\hat{\sigma}$ is the median absolute deviation defined by:

$$\hat{\sigma} = 1.4826 \times M(|\xi_{it} - M(\xi_{it})|).$$

where M is the median. The M estimator of β based on the function $\rho(\xi_{it})$. Differentiating the objective function ρ with respect to the coefficients β , and setting the partial derivatives to zero in (7), we find:

$$\min_{\beta} \sum_{i=1}^N \sum_{t=1}^T \psi \left(\frac{\xi_{it}}{\hat{\sigma}} \right) \frac{\partial}{\partial \beta} \xi_{it} = 0 \quad (8)$$

where $\psi(u_{it}) = \rho'(u_{it})$ is the derivative for $\rho(u_{it})$, where $\psi(u_{it})$ is called the influence (score) function. If we define a weight function $W_{\xi}(u_{it}) = \frac{\psi(u_{it})}{u_{it}}$, we can rewrite (8) by using the weight, and then, we can obtain the first-order condition for M estimators as following:

$$\sum_{i=1}^N \sum_{t=1}^T \xi'_{it} W_{\xi}(\xi_{it}) \xi_{it} = 0 \quad (9)$$

where $\xi'_{it} = \frac{\partial}{\partial \beta} \xi_{it}$ and $W_{\xi}(\xi_{it})$ is an $(NT \times NT)$ matrix. When solving (9) depending on the weights of Huber, Hampel, and Tukey bisquare, we can obtain three robust estimators for β in the FEP model. We can write the weight function of Huber, Hampel, and Tukey bisquare in panel data regression as follows [10]:

- Huber's weight function:

$$W_{HR}(\xi_{it}) = \begin{cases} 1, & \text{if } |u_{it}| < 1.345 \\ \frac{1.345}{|u_{it}|}, & \text{if } |u_{it}| \geq 1.345 \end{cases}$$

- Hampel's weight function:

$$W_{HM}(\xi_{it}) = \begin{cases} 1, & \text{if } |u_{it}| < 2 \\ \frac{2}{|u_{it}|}, & \text{if } 2 \leq |u_{it}| < 4 \\ 2 \left[\frac{8/|u_{it}| - 1}{8 - 4} \right], & \text{if } 4 \leq |u_{it}| \leq 8 \\ 0, & \text{otherwise} \end{cases}$$

- Tukey's weight function:

$$W_{TK}(\xi_{it}) = \begin{cases} [1 - (u_{it}/4.685)^2]^2, & \text{if } |u_{it}| \leq 4.685 \\ 0, & \text{if } |u_{it}| > 4.685 \end{cases}$$

We will show the algorithm of the robust M estimation method for the FEP model, which should be used when the count panel data contain outliers to obtain a robust estimation.

1. Estimating regression coefficients for the count panel data by using the FEP model.
2. Calculate initial parameters ($\hat{\beta}_{FPML}$) by using ML estimation method.
3. Calculate residuals value $\xi_{it} = \frac{R_{it}(\alpha_i, \beta)}{\sqrt{\hat{\alpha}_i \lambda_{it}}}$.
4. Calculate median absolute deviation $\hat{\sigma}_i$.
5. Calculate standardized residuals $u_{it} = \frac{\xi_{it}}{\hat{\sigma}_i}$.
6. Calculate the weight value $W_{\xi}(\xi_{it})$.
7. Estimate $\hat{\beta}_{FPHR}$, $\hat{\beta}_{FPHM}$, and $\hat{\beta}_{FPTK}$ estimators based on $W_{\xi}(\xi_{it})$.
8. Repeat steps 3-6 to obtain a convergent value of $\hat{\beta}_{FPHR}$, $\hat{\beta}_{FPHM}$, and $\hat{\beta}_{FPTK}$ estimators.
9. Examine the significance of the independent variables on the dependent variable and compare the performance of these estimators using some criteria, for example, use:

$$\text{AIC} = 2p - 2 \log L; \quad \text{BIC} = p \log(n) - 2 \log L.$$

where p is the number of parameters, n denotes the total number of observations, and L denotes ML value.

It is known that Classical estimation methods are affected by outliers like ordinary least squares and ML. Therefore, robust estimation methods have been proposed to deal with outliers to obtain good and stable results in the presence of outliers. To achieve this resistance, we used the M estimation method to achieve high robustness or high efficiency for estimators, where the M estimator is more efficient than traditional estimators, see, e.g. [41, 31].

4. Monte Carlo Simulation Study

The Monte Carlo simulation was conducted to examine the effect of outliers on estimates in the FEP model. Some studies have been relied upon when designing simulation, see [40, 2]. R software is used to conduct our simulation.

4.1. Algorithm of Simulation

The algorithm of the Monte Carlo simulation study for the FEP model is based on the following:

1. We design the panel data set to obtain the total number of observations ($n = N \times T$) with the following steps:
 - (a) The values of cross-section (N) were chosen to be 20, 50, 100, and 200 to represent small, moderate, and large samples for the number of individuals, respectively.
 - (b) The values of Time-series (T) were chosen to be 5, 10, and 20 to represent different sizes for the time period.
2. We generate count panel data as follows:
 - (a) The vector of true parameters was chosen to be $\beta = 1$.
 - (b) The independent variables ($X_{k,it}$) were generated from uniform distribution on interval $(-0.5, 1)$, where the number of independent variables was $k = 3$ and 6.
 - (c) The dependent variable (y_{it}) was generated from the Poisson distribution with a mean equal to $\exp(\delta_i + X'_{k,it}\beta)$.
 - (d) The percentage of outliers ($\tau\%$) in the dependent variable was chosen to be 0, 5, 10, and 20. This percentage is calculated from the total number of observations (n). When the proportion of outliers equals zero ($\tau = 0\%$) this means that count panel data do not contain outliers.
 - (e) The outliers generated from Poisson distribution with mean equal to $8 \text{ IQR}[\exp(\delta_i + X'_{k,it}\beta)]$, where IQR is the interquartile range.
3. Estimating regression coefficients with the following steps:
 - (a) Estimate parameters ($\hat{\beta}_{FPMML}$) by using ML estimation method.
 - (b) Calculate residuals value $\xi_{it} = \frac{R_{it}(\alpha_i, \beta)}{\sqrt{\hat{\alpha}_i \lambda_{it}}}$
 - (c) Calculate median absolute deviation $\hat{\sigma}_i$.
 - (d) Calculate standardized residuals $u_{it} = \frac{\xi_{it}}{\hat{\sigma}_i}$.
 - (e) Calculate the value of weights $W_{HR}(\xi_{it})$, $W_{HM}(\xi_{it})$, and $W_{TK}(\xi_{it})$.
 - (f) Estimate regression coefficients: FPHR estimator ($\hat{\beta}_{FPHR}$), FPHM estimator ($\hat{\beta}_{FPHM}$), and FPTK estimator ($\hat{\beta}_{FPTK}$) estimator using weighted ML estimation method with weight $W_\xi(\xi_{it})$.
4. For all experiments of simulation, we ran 1000 replications.
5. We examine the performance of these estimators with the following steps:
 - (a) Calculate the mean squared error (MSE) and mean absolute error (MAE) for N, T, P , and τ different for each parameter separately as follows:

$$\text{MSE} = \frac{1}{1000} \sum_{l=1}^{1000} (\hat{\beta}_l - \beta)^2; \quad \text{MAE} = \frac{1}{1000} \sum_{l=1}^{1000} |\hat{\beta}_l - \beta|.$$

where $\hat{\beta}_l$ is the vector of estimated values of β at l^{th} experiment of 1000 replication of simulation, while β is the vector of true parameters, where better estimator is the one has small total MSE and total MAE.

- (b) Calculate the mean relative efficiency (MRE) of M estimators for all T and τ for each N separately to compare the performance of estimators (FPHR, FPHM, and FPTK). The MRE is calculated as:

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

where $\hat{\beta}_R$ represents $\hat{\beta}_{FPHR}$, $\hat{\beta}_{FPHM}$, and $\hat{\beta}_{FPTK}$ estimators. The best efficient estimator is the one with the largest MRE.

4.2. Simulation Results

The results of the Monte Carlo simulation for the small, moderate, and large samples have been provided in Tables 1 to 4. Specifically, Tables 1, 2, 3, and 4 presented the total MSE and the total MAE values of all estimators (non-robust and robust) when the number of explanatory variables equal 3 and 6, with the variation in cross-section sizes to be 20, 50, 100, and 200, and the time-series periods to be 5, 10, and 20.

According to the results of these tables, the values of MSE and MAE are increasing when both the number of explanatory variables and the percentage of outliers increase. When the proportion of outliers increases from 5% to 20%, the MSE and MAE values are inflated. But, this increase is somehow large for the FPML estimator and small for the FPHR, FPHM, and FPTK estimators. In the case of count panel data that do not contain outliers ($\tau = 0\%$), the non-robust estimator (FPML) performs better than robust estimators (FPHR, FPHM, and FPTK) for all values of N and T .

On the other hand, in general, the MSE and MAE values of estimators decrease when the cross-section and time-series increase. Where robust estimators (FPHR, FPHM, and FPTK) are better than the non-robust estimator (FPML) when the count panel data contain outliers for all different values of N and T . The rate of decrease for MSE and MAE values of robust estimators is greater compared to the non-robust estimator.

Figures 1 and 2 show the MRE for robust estimators (FPHR, FPHM, and FPTK) clustered by time-series (from 5 to 20) and percentages of outliers (from 5% to 20%) for each cross-section separately, when the number of parameters ($k = 3$ and 6).

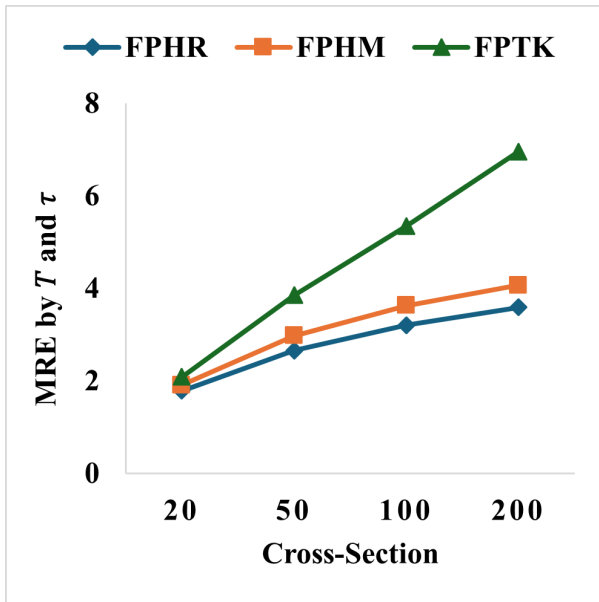


Figure 1. The MRE of the Robust Estimators when $k = 3$.

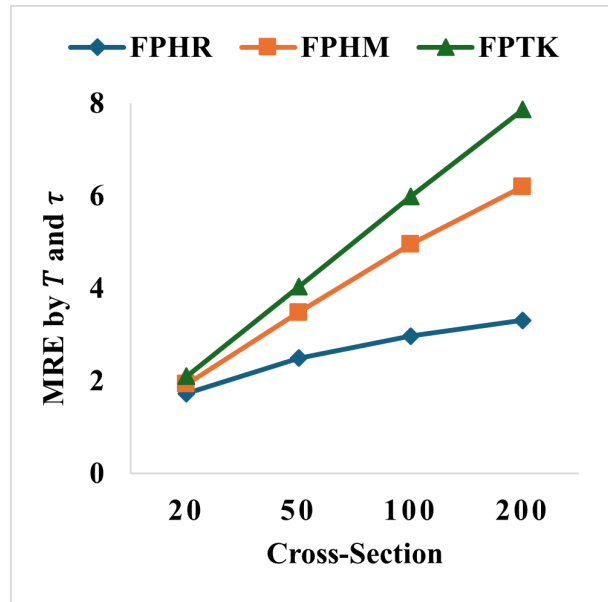


Figure 2. The MRE of the Robust Estimators when $k = 6$.

These figures indicate that the MRE values of the FPTK estimator are larger than the MRE values of FPHR and FPHM estimators for each cross-section value, this means that the FPTK estimator is more efficient than the FPHR and FPHM estimators in different N , T , and τ values. When N increases for each T and τ , the efficiency of FPTK increases. In Figure 2, the efficiency of the FPHM increases, but the FPTK estimator is still more efficient than the FPHM estimator. Efficiency increases when $N = 200$, $T = 20$, and $\tau = 20\%$.

Table 1. MSE and MAE Values of Estimators when $N = 20$.

T		$k = 3$				$k = 6$			
		FPML	FPHM	FPHR	FPTK	FPML	FPHM	FPHR	FPTK
$\tau = 0\%$									
5	MSE	0.5798	0.7753	0.8177	0.9552	1.2305	1.6470	1.6948	1.9952
	MAE	1.0413	1.1975	1.2259	1.3186	2.1466	2.4806	2.5137	2.7278
10	MSE	0.2299	0.2980	0.3447	0.3904	0.5349	0.6492	0.6918	0.8117
	MAE	0.6597	0.7496	0.8029	0.8496	1.4208	1.5696	1.6148	1.7476
20	MSE	0.0997	0.1242	0.1554	0.1705	0.2508	0.2891	0.3295	0.3817
	MAE	0.4333	0.4882	0.5411	0.5658	0.9671	1.0467	1.1158	1.2009
$\tau = 5\%$									
5	MSE	0.8103	0.7057	0.7243	0.7816	2.2280	1.6471	1.7231	1.7798
	MAE	1.2647	1.1411	1.1625	1.2054	2.9434	2.4810	2.5403	2.5809
10	MSE	0.4415	0.2815	0.2934	0.3072	1.0587	0.6090	0.6393	0.6474
	MAE	0.9418	0.7369	0.7513	0.7690	2.0923	1.5223	1.5608	1.5719
20	MSE	0.3335	0.1285	0.1315	0.1388	0.8037	0.2855	0.3091	0.3092
	MAE	0.8616	0.4955	0.5012	0.5165	1.8877	1.0444	1.0796	1.0854
$\tau = 10\%$									
5	MSE	1.1827	0.7768	0.8191	0.7873	2.9585	1.8420	1.9511	1.7877
	MAE	1.5414	1.2108	1.2428	1.2245	3.5191	2.6222	2.7224	2.5887
10	MSE	0.7683	0.3286	0.3502	0.3221	1.8812	0.7623	0.8845	0.7382
	MAE	1.3312	0.7915	0.8173	0.7815	2.9313	1.7016	1.8544	1.6808
20	MSE	0.6750	0.1750	0.2244	0.1783	1.5953	0.3520	0.5108	0.3724
	MAE	1.3043	0.5755	0.6588	0.5826	2.8619	1.1587	1.4255	1.1985
$\tau = 15\%$									
5	MSE	1.4485	0.9745	0.9839	0.8952	3.5728	2.4803	2.5432	2.2121
	MAE	1.7916	1.3607	1.3775	1.3042	3.9985	3.0688	3.1420	2.8937
10	MSE	0.9887	0.4339	0.4678	0.3624	2.7666	0.9209	1.2691	0.7943
	MAE	1.5611	0.9221	0.9653	0.8365	3.7459	1.8762	2.2699	1.7491
20	MSE	1.0594	0.2518	0.3585	0.2040	1.7995	0.5361	0.7058	0.4398
	MAE	1.7048	0.7016	0.8742	0.6280	3.1093	1.4665	1.7447	1.3132
$\tau = 20\%$									
5	MSE	1.6245	1.1568	1.1520	1.0321	3.9201	3.0191	3.0351	2.7369
	MAE	1.9318	1.4879	1.4865	1.3865	4.2840	3.4117	3.4608	3.2206
10	MSE	1.4559	0.6799	0.7482	0.5142	2.8165	1.3773	1.5721	1.0837
	MAE	1.9660	1.1775	1.2624	1.0000	3.8420	2.3360	2.5729	2.0283
20	MSE	1.3470	0.4721	0.5659	0.3025	2.9461	0.8807	1.3284	0.6138
	MAE	1.9484	1.0084	1.1457	0.7727	4.0820	1.9286	2.5241	1.5691

Table 2. MSE and MAE Values of Estimators when $N = 50$.

T		$k = 3$				$k = 6$			
		FPML	FPHM	FPHR	FPTK	FPML	FPHM	FPHR	FPTK
$\tau = 0\%$									
5	MSE	0.1984	0.2418	0.2854	0.3133	0.4499	0.5615	0.6172	0.7037
	MAE	0.6139	0.6784	0.7369	0.7676	1.3009	1.4528	1.5186	1.6247
10	MSE	0.0879	0.1074	0.1374	0.1499	0.2212	0.2417	0.2827	0.3225
	MAE	0.4094	0.4520	0.5130	0.5344	0.9166	0.9617	1.0372	1.1041
20	MSE	0.0427	0.0522	0.0765	0.0785	0.1028	0.1142	0.1463	0.1667
	MAE	0.2846	0.3158	0.3825	0.3850	0.6254	0.6606	0.7484	0.7971
$\tau = 5\%$									
5	MSE	0.4196	0.2476	0.2568	0.2740	1.0666	0.5375	0.5682	0.5868
	MAE	0.9281	0.6830	0.6973	0.7164	2.0973	1.4289	1.4683	1.4963
10	MSE	0.3209	0.1188	0.1184	0.1251	0.6133	0.2553	0.2621	0.2680
	MAE	0.8483	0.4765	0.4760	0.4900	1.6344	0.9925	1.0019	1.0132
20	MSE	0.2535	0.0611	0.0572	0.0601	0.6704	0.1363	0.1483	0.1422
	MAE	0.7970	0.3445	0.3314	0.3404	1.8501	0.7270	0.7588	0.7410
$\tau = 10\%$									
5	MSE	0.7926	0.3101	0.3412	0.3024	1.8817	0.6374	0.7916	0.6506
	MAE	1.3603	0.7651	0.8043	0.7545	2.9716	1.5556	1.7457	1.5750
10	MSE	0.6778	0.1437	0.1907	0.1453	1.5688	0.2977	0.4560	0.3195
	MAE	1.3291	0.5272	0.6180	0.5318	2.8781	1.0663	1.3557	1.1078
20	MSE	0.6298	0.0852	0.1284	0.0855	1.1855	0.1892	0.2989	0.1903
	MAE	1.3290	0.4090	0.5196	0.4095	2.5512	0.8678	1.1372	0.8718
$\tau = 15\%$									
5	MSE	1.1143	0.4260	0.4699	0.3462	2.5503	0.8296	1.1151	0.7214
	MAE	1.6840	0.9152	0.9718	0.8132	3.6239	1.7791	2.1316	1.6645
10	MSE	1.0054	0.2553	0.3254	0.1868	2.1158	0.4491	0.7415	0.3948
	MAE	1.6685	0.7219	0.8428	0.6046	3.4212	1.3273	1.8085	1.2417
20	MSE	0.9242	0.1617	0.2512	0.1150	1.8667	0.3138	0.5841	0.2653
	MAE	1.6328	0.5886	0.7791	0.4825	3.2731	1.1515	1.6896	1.0488
$\tau = 20\%$									
5	MSE	1.4183	0.6922	0.6959	0.4832	3.0953	1.2952	1.6005	0.9734
	MAE	1.9507	1.2131	1.2239	0.9732	4.0716	2.2542	2.6155	1.9223
10	MSE	1.1909	0.4749	0.4795	0.2772	2.6454	0.8391	1.1711	0.5713
	MAE	1.8418	1.0637	1.0766	0.7620	3.8688	1.9107	2.3732	1.5222
20	MSE	1.2558	0.3867	0.4577	0.1994	2.7385	0.5518	1.1003	0.3710
	MAE	1.9170	0.9905	1.1016	0.6663	4.0013	1.5869	2.4231	1.2672

Table 3. MSE and MAE Values of Estimators when $N = 100$.

T		$k = 3$				$k = 6$			
		FPML	FPHM	FPHR	FPTK	FPML	FPHM	FPHR	FPTK
$\tau = 0\%$									
5	MSE	0.1014	0.1306	0.1599	0.1756	0.2350	0.2759	0.3242	0.3690
	MAE	0.4367	0.4960	0.5499	0.5750	0.9431	1.0229	1.1096	1.1790
10	MSE	0.0400	0.0510	0.0735	0.0767	0.1120	0.1241	0.1568	0.1789
	MAE	0.2773	0.3107	0.3737	0.3804	0.6514	0.6878	0.7745	0.8271
20	MSE	0.0194	0.0245	0.0450	0.0436	0.0541	0.0584	0.0865	0.1018
	MAE	0.1943	0.2170	0.2984	0.2901	0.4532	0.4726	0.5792	0.6281
$\tau = 5\%$									
5	MSE	0.2971	0.1357	0.1343	0.1445	0.7666	0.2704	0.2916	0.2909
	MAE	0.8077	0.5147	0.5114	0.5278	1.8468	1.0135	1.0540	1.0538
10	MSE	0.2449	0.0622	0.0586	0.0629	0.6147	0.1411	0.1537	0.1494
	MAE	0.7787	0.3458	0.3346	0.3466	1.7363	0.7404	0.7718	0.7596
20	MSE	0.2447	0.0352	0.0317	0.0336	0.5631	0.0851	0.0910	0.0824
	MAE	0.8175	0.2615	0.2450	0.2531	1.7506	0.5823	0.6018	0.5684
$\tau = 10\%$									
5	MSE	0.6102	0.1754	0.1920	0.1614	1.5098	0.3355	0.4737	0.3501
	MAE	1.2414	0.5823	0.6118	0.5550	2.7865	1.1357	1.3710	1.1614
10	MSE	0.6168	0.0932	0.1297	0.0883	1.2821	0.1935	0.3136	0.2008
	MAE	1.3130	0.4322	0.5206	0.4173	2.6617	0.8755	1.1629	0.8953
20	MSE	0.5714	0.0648	0.0973	0.0585	1.3364	0.1286	0.2716	0.1485
	MAE	1.2871	0.3726	0.4749	0.3487	2.7788	0.7348	1.1463	0.7999
$\tau = 15\%$									
5	MSE	0.8902	0.2764	0.2992	0.1996	2.2230	0.5100	0.8110	0.4425
	MAE	1.5572	0.7546	0.7962	0.6243	3.4894	1.4140	1.8840	1.3187
10	MSE	1.0005	0.1725	0.2715	0.1220	2.1219	0.3083	0.6454	0.2781
	MAE	1.6991	0.6068	0.8095	0.4971	3.4965	1.1300	1.7838	1.0726
20	MSE	0.9287	0.1465	0.2342	0.0952	1.9320	0.2388	0.5557	0.2098
	MAE	1.6531	0.5948	0.7890	0.4612	3.3688	1.0493	1.7307	0.9802
$\tau = 20\%$									
5	MSE	1.2038	0.5302	0.5045	0.3218	2.7721	0.8529	1.2223	0.5951
	MAE	1.8399	1.1151	1.0856	0.8109	3.9532	1.8924	2.4035	1.5462
10	MSE	1.1400	0.4200	0.4278	0.2214	2.6668	0.6182	1.0568	0.3933
	MAE	1.8225	1.0439	1.0599	0.7100	3.9449	1.7004	2.3656	1.3033
20	MSE	1.2459	0.3702	0.4323	0.1736	2.7922	0.4737	1.0498	0.2973
	MAE	1.9211	1.0086	1.1022	0.6552	4.0677	1.5466	2.4357	1.1857

Table 4. MSE and MAE Values of Estimators when $N = 200$.

T		$k = 3$				$k = 6$			
		FPML	FPHM	FPHR	FPTK	FPML	FPHM	FPHR	FPTK
$\tau = 0\%$									
5	MSE	0.0507	0.0604	0.0886	0.0905	0.1177	0.1273	0.1623	0.1837
	MAE	0.3112	0.3390	0.4124	0.4149	0.6683	0.6966	0.7890	0.8390
10	MSE	0.0220	0.0259	0.0493	0.0479	0.0547	0.0608	0.0949	0.1073
	MAE	0.2030	0.2224	0.3132	0.3049	0.4564	0.4848	0.6084	0.6472
20	MSE	0.0106	0.0128	0.0344	0.0307	0.0284	0.0298	0.0604	0.0698
	MAE	0.1424	0.1574	0.2703	0.2494	0.3284	0.3385	0.4953	0.5324
$\tau = 5\%$									
5	MSE	0.2222	0.0712	0.0642	0.0690	0.5754	0.1490	0.1546	0.1539
	MAE	0.7290	0.3721	0.3510	0.3642	1.6680	0.7622	0.7748	0.7739
10	MSE	0.2217	0.0361	0.0303	0.0322	0.5665	0.0828	0.0887	0.0828
	MAE	0.7748	0.2657	0.2401	0.2473	1.7535	0.5755	0.5941	0.5721
20	MSE	0.2336	0.0254	0.0202	0.0218	0.5870	0.0599	0.0672	0.0590
	MAE	0.8167	0.2264	0.1980	0.2058	1.8304	0.5005	0.5313	0.4923
$\tau = 10\%$									
5	MSE	0.5153	0.1101	0.1209	0.0920	1.2956	0.1994	0.3205	0.2107
	MAE	1.1858	0.4695	0.4965	0.4238	2.6723	0.8884	1.1729	0.9168
10	MSE	0.5240	0.0748	0.0918	0.0581	1.1483	0.1326	0.2397	0.1348
	MAE	1.2294	0.4028	0.4559	0.3448	2.5697	0.7465	1.0669	0.7519
20	MSE	0.5683	0.0479	0.0842	0.0450	1.2984	0.0972	0.2401	0.1184
	MAE	1.2942	0.3337	0.4645	0.3206	2.7670	0.6681	1.1326	0.7525
$\tau = 15\%$									
5	MSE	0.8257	0.2319	0.2414	0.1382	1.9741	0.3291	0.6105	0.2827
	MAE	1.5344	0.7299	0.7501	0.5330	3.3638	1.1680	1.7191	1.0767
10	MSE	0.9035	0.1627	0.2306	0.0980	1.9786	0.2302	0.5518	0.2029
	MAE	1.6292	0.6331	0.7804	0.4703	3.4087	1.0242	1.7205	0.9590
20	MSE	0.9356	0.1283	0.2212	0.0791	2.0499	0.1853	0.5544	0.1788
	MAE	1.6668	0.5823	0.7886	0.4430	3.4903	0.9649	1.7767	0.9525
$\tau = 20\%$									
5	MSE	1.1101	0.4742	0.4261	0.2516	2.5673	0.6716	1.0323	0.4262
	MAE	1.7955	1.1161	1.0511	0.7567	3.8645	1.7785	2.3232	1.3572
10	MSE	1.2217	0.3803	0.4250	0.1779	2.5312	0.5357	0.9705	0.3279
	MAE	1.9014	1.0234	1.0911	0.6645	3.8686	1.6621	2.3336	1.2544
20	MSE	1.2124	0.3488	0.4134	0.1583	2.6468	0.4293	0.9816	0.2675
	MAE	1.9012	1.0010	1.0957	0.6550	3.9713	1.5311	2.3880	1.1833

5. Empirical Application

In this application, we analyzed the patent applications for 18 high-income countries in Europe over the period from 2005 to 2020, these countries investing substantial amounts in research and development (R&D). The count panel data was obtained from the World Bank website. In our study, the response variable (y_{it}) is the number of patent applications (count panel), while the independent variables include: The number of researchers in R&D per 1000 people ($X_{1,it}$), the logarithm of the R&D expenditures ($X_{2,it}$), the logarithm of the gross domestic product per capita ($X_{3,it}$), the logarithm of the information and communication technology goods imports ($X_{4,it}$), the logarithm of the information and communication technology goods exports ($X_{5,it}$), and unemployment rate ($X_{6,it}$) which is a percentage of the total labor force.

5.1. Application Results

Table 5 presents some descriptive statistics of the variables selected in our study. In this application, we will use the algorithm described in Section 3 to obtain the results of the estimators.

Table 5. Descriptive Statistics of the Variables

Variables	Mean	Median	Min.	Max.	Std. Dev.
y_{it}	1748.35	786.00	15.00	14748.00	3189.38
$X_{1,it}$	4.05	3.74	1.31	7.93	1.73
$X_{2,it}$	26.51	26.91	22.92	29.50	1.67
$X_{3,it}$	10.35	10.47	8.94	11.73	0.67
$X_{4,it}$	27.18	27.42	24.57	29.55	1.26
$X_{5,it}$	26.69	26.89	22.96	29.54	1.48
$X_{6,it}$	0.08	0.07	0.02	0.26	0.04

Figure 3 shows a box plot of patents for countries under study. This figure shows that the count panel data for the response variable contains outlier values.

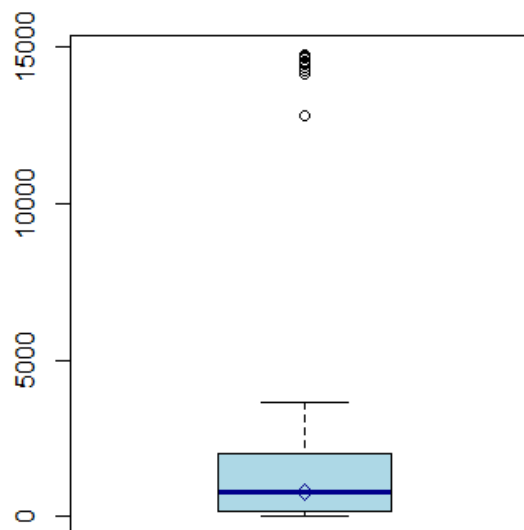


Figure 3. The Box plot of the Number of Patents.

Table 6 presents the results of non-robust and robust estimates for the FEP model. It can be seen that all the explanatory variables for FPHM, FPHR, and FPTK estimators have statistically significant effects on the response variable, so the estimated coefficients are suitable for the robust FEP model. In the non-robust FEP model, it can be noted that all the explanatory variables are statistically significant except $X_{4,it}$.

Table 6. Estimates of Fixed Effects Poisson Panel Data Model

Variables	FPML		FPHM		FPHR		FPTK	
	Estimate	Std. E.	Estimate	Std. E.	Estimate	Std. E.	Estimate	Std. E.
Intercept	-18.425*	0.115	-9.184*	0.139	-11.035*	0.137	-8.962*	0.145
$X_{1,it}$	-0.979*	0.006	-0.570*	0.008	-0.664*	0.008	-0.524*	0.009
$X_{2,it}$	1.075*	0.005	0.660*	0.005	0.733*	0.005	0.651*	0.005
$X_{3,it}$	-0.243*	0.006	-0.085*	0.007	-0.109*	0.006	-0.072*	0.007
$X_{4,it}$	-0.004	0.003	-0.022*	0.004	-0.008**	0.004	-0.026*	0.004
$X_{5,it}$	0.032*	0.002	0.040*	0.003	0.035*	0.003	0.040*	0.003
$X_{6,it}$	1.994*	0.166	10.467*	0.292	6.426*	0.272	12.746*	0.314

Note: *, ** indicate that the level of significance is at 1% and 5%, respectively.

Based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) presented in Figure 4, we concluded that the robust estimators FPHM, FPHR, and FPTK performed better than the non-robust estimator FPML, where the best estimator is the estimator that has a minimum value of AIC or BIC.

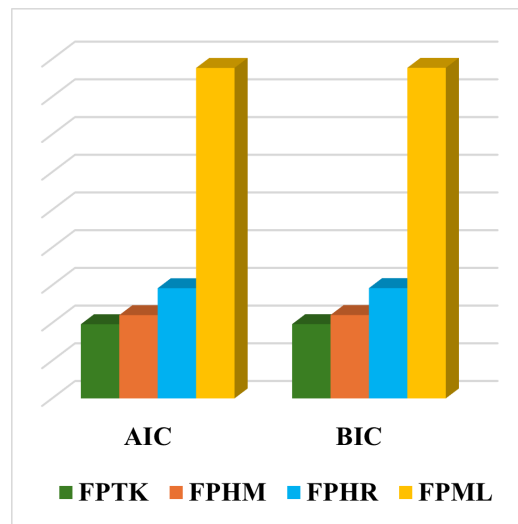


Figure 4. The AIC and BIC Values for Estimators

6. Conclusions

This paper presented a robust estimation method for the FEP model for analyzing count panel data with outliers. We have introduced robust estimators based on the M estimation method, including FPHM, FPHR, and FPTK, and compared these with the non-robust FPML estimator.

To examine the performance of the estimators, we conducted a Monte Carlo simulation study and a practical empirical application to patent data from high-income European countries. The simulation result shows that, in

case there are no outliers, the FPML estimator is better, while the robust estimators have lower performance. In the presence of outliers, the weighted ML estimators (FPHM, FPHR, and FPTK) are more effective compared to the ML estimator. The result of the application shows that robust estimators are better than non-robust estimators with outliers. In addition, FPTK is more efficient than FPHM and FPHR.

This study is expected to provide useful information for both researchers and policy makers involved in scientific research and development. Furthermore, it improves the statistical methods used to analyze the count panel data, particularly when dealing with outliers.

REFERENCES

1. M.R. Abonazel, *Generalized Estimators of Stationary Random-Coefficients Panel Data Models: Asymptotic and Small Sample Properties*, Revstat Statistical Journal, vol. 17, no. 4, pp. 493–521, 2019.
2. M.R. Abonazel and O. M. Saber, *A Comparative Study of Robust Estimators for Poisson Regression Model with Outliers*, Journal of Statistics Applications & Probability, vol. 9, no. 2, pp. 279–286, 2020.
3. M.R. Abonazel, S.M. El-sayed, and O.M. Saber, *Performance of robust count regression estimators in the case of overdispersion, zero inflated, and outliers: simulation study and application to German health data*, Communications in Mathematical Biology and Neuroscience, vol. 2021, 2021.
4. S.O. Adams, M.A. Bamanga, S.O. Olanrewaju, H.U. Yahaya, and R.O. Akano, *Modeling COVID-19 Cases in Nigeria Using Some Selected Count Data Regression Models*, International Journal of Healthcare and Medical Sciences, vol. 6, no. 4, pp. 64–73, 2020.
5. P.D. Allison and R.P. Waterman, *Fixed-Effects Negative Binomial Regression Models*, Sociological Methodology, vol. 32, no. 1, pp. 247–265, 2002.
6. M. Amelia, K. Sadik, and B. Sartono, *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.
7. J.F. Angers, D. Desjardins, G. Dionne, and F. Guertin, *Modelling and Estimating Individual and Firm Effects with Count Panel Data*, The Journal of the International Actuarial Association, vol. 48, no. 3, pp. 1049–1078, 2018.
8. E. AVCI, *Flexibility of Using Com-Poisson Regression Model for Count Data*, Statistics, Optimization & Information Computing, vol. 6, no. 2, pp. 278–285, 2018. <https://doi.org/10.19139/soic.v6i2.278>
9. B.H. Baltagi, *Econometric Analysis of Panel Data*, John Wiley and Sons Ltd, England, 2005.
10. B.H. Baltagi, *The Oxford Handbook of Panel Data*, Oxford University Press, United States of America, 2015.
11. W. Bari and B.C. Sutradhar, *Robust Inferences in Longitudinal Models for Binary and Count Panel Data in the Presence of Outliers*, The Indian Journal of Statistics, Series B, vol. 72, pp. 11–37, 2010.
12. B.H. Beyaztas and S. Bandyopadhyay, *Data Driven Robust Estimation Methods for Fixed Effects Panel Data Models*, Journal of Statistical Computation and Simulation, vol. 92, no. 7 pp. 1401–1425, 2022.
13. M.C. Bramati and C. Croux, *Robust Estimators for the Fixed Effects Panel Data Model*, The Econometrics Journal, vol. 10, no. 3, pp. 521–540, 2007.
14. M. Cincera, *Patents, R&D, and Technological Spillovers at the Firm Level: Some Evidence from Econometric Count Models for Panel Data*, Journal of Applied Econometrics, vol. 12, no. 3, pp. 265–280, 1997.
15. L. Fridstrøm, J. Ifver, S. Ingebrigtsen, R. Kulmala, and L.K. Thomsen, *Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts*, Accident Analysis & Prevention, vol. 27, no. 1, pp. 1–20, 1995.
16. P. Guimaraes, *The Fixed Effects Negative Binomial Model Revisited*, Economics Letters, vol. 99, no. 1, pp. 63–66, 2008.
17. S. Gurmu and F. Pérez-Sebastián, *Patents, R&D and Lag Effects: Evidence from Flexible Methods for Count Panel Data on Manufacturing Firms*, Empirical Economics, vol. 35, no. 3, pp. 507–526, 2008.
18. J.A. Hausman, B.H. Hall, and Z. Griliches, *Econometric Models for Count Data with An Application to the patents-R&D Relationship*, Econometrica, vol. 52, no. 4, pp. 909–938, 1984.
19. P. Huber, *Robust Estimation of A Location Parameter*, The Annals of Mathematical Statistics, vol. 35, no. 1, pp. 73–101, 1964.
20. I.A. Idriss and W.H. Cheng, *Robust Estimators for Poisson Regression*, Open Journal of Statistics, vol. 13, no. 1, pp. 112–118, 2023.
21. N. Ismail and H. Zamani, *Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models*, In Casualty Actuarial Society E-Forum, vol. 41, no. 20, pp. 1–28, 2013.
22. M.G. Karlaftis and A.P. Tarko, *Heterogeneity Considerations in Accident Modeling*, Accident Analysis & Prevention, vol. 30, no. 4, pp. 425–433, 1998.
23. S.S.P. Kumara and H.C. Chin, *Study of Fatal Traffic Accidents in Asia Pacific Countries*, Transportation Research Record, no. 1897, pp. 43–47, 2004.
24. Q. Li, H. Chen, and F. Zhu, *Robust estimation for Poisson integer-valued GARCH models using a new hybrid loss*, Journal of Systems Science and Complexity, vol. 34, pp. 1578–1596, 2021.
25. A.F. Lukman, M. Arashi, and V. Prokaj, *Robust biased estimators for Poisson regression model: simulation and applications*, Concurrency Computation Practice and Experience, vol. 35, no. 7, e7594, 2023.
26. R.B. Noland and M.A. Quddus, *Analysis of Pedestrian and Bicycle Casualties with Regional Panel Data*, Transportation Research Record, no. 1897, pp. 28–33, 2004.
27. N. Pillai, *Panel data analysis with stata part I fixed effects and random effects models*, University Library of Munich, Germany, 2016.
28. X. Piulachs, R. Alemany, M. Guillén, and D. Rizopoulos, *Joint Models for Longitudinal Counts and Left-Truncated Time-To Event Data with Applications to Health Insurance*, SORT-Statistics and Operations Research Transactions, vol. 41, no. 2, pp. 347–372,

- 2017.
29. A.E. Renshaw, *Modelling the Claims Process in the Presence of Covariates*, ASTIN Bulletin: The Journal of the IAA, vol. 24, no. 2, pp. 265–285, 1994.
 30. R.T. Riphahn, A. Wambach, and A. Million, *Incentive Effects in The Demand for Health Care: A Bivariate Panel Count Data Estimation*, Journal of Applied Econometrics, vol. 18, no. 4, pp. 387–405, 2003.
 31. P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, 1987.
 32. S. Russo, D. Flender, and G.F. da Silva, *Poisson Regression Models for Count Data: Use in the Number of Deaths in the Santo Angelo (Brazil)*, Journal of Basic & Applied Sciences, vol. 8, no. 2, pp. 266–269, 2012.
 33. J. Á. Vřšek, *Estimating the Model with Fixed and Random Effects by A Robust Method*, Methodology and Computing in Applied Probability, vol. 17, no. 4, pp. 999–1014, 2015.
 34. R. Winkelmann, *Econometric Analysis of Count Data*, Springer-Verlag Berlin Heidelberg, 2008.
 35. W.F.W. Yaacob, M.A. Lazim, and Y.B. Wah, *A Practical Approach in Modelling Count Data*, In Proceedings of the Regional Conference on Statistical Sciences, Malaysia Institute of Statistics, Faculty of Computer and Mathematical Sciences, University Technology MARA (UiTM), Malaysia, pp. 176–183, 2010.
 36. W.F.W. Yaacob, M.A. Lazim, and Y.B. Wah, *Applying Fixed Effects Panel Count Model to Examine Road Accident Occurrence*, Journal of Applied Sciences, vol. 11, no. 7, pp. 1185–1191, 2011.
 37. W.F.W. Yaacob, M.A. Lazim, and Y.B. Wah, *Modeling Road Accidents using Fixed Effects Model: Conditional versus Unconditional Model*, Proceedings of the World Congress on Engineering, vol. 1, pp. 218–223, 2012.
 38. A.H. Youssef and M.R. Abonazel, *A Comparative Study for Estimation Parameters in Panel Data Model*, InterStat Journal, May 2009, no. 2, pp. 1–17, 2009.
 39. A.H. Youssef, M.R. Abonazel, and E.G. Ahmed, *Estimating the Number of Patents in the World Using Count Panel Data Models*, Asian Journal of Probability and Statistics, vol. 6, no. 4, pp. 24–33, 2020.
 40. A.H. Youssef, M.R. Abonazel, and E.G. Ahmed, *The Performance of Count Panel Data Estimators: A Simulation Study and Application to Patents in Arab Countries*, Journal of Mathematical and Computational Science, vol. 11, no. 6, pp. 8173–8196, 2021.
 41. A.H. Youssef, M.R. Abonazel, and A.R. Kamel, *Efficiency Comparisons of Robust and Non-Robust Estimators for Seemingly Unrelated Regressions Model*, WSEAS Transactions on Mathematics, vol. 21, pp. 218–244, 2022.