AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

# Comparison of Estimation Methods for Random Coefficient Panel Data Model

https://doi.org/10.29124/kjeas.1651.14

**Hassan Hopoop Razaq** (1)

hasanhopop@utq.edu.iq

Department of Economics. College of Administration and Economics .University of Thi-Qar

**. Ahmed Abed Naeem** (2)

anaeem@uowasit.edu.iq

Department of Administration Business. College of Administration and Economics .University of Wasit.

## Abstract

Panel data models are among the commonly used models in many fields. In this research we relied on the random coefficients model, which is characterized by the fact that the marginal tendencies are random and not fixed, meaning that they have a mean and a covariance matrix . The mean parameter is estimated by three methods are, two-steps Aitken estimator (2SAE) , fixed effect estimator (FE) and group mean estimator (GME). Using Mean of Means Absolute Error to compare between estimation methods, the simulation results showed that the (2SAE) estimator is the best estimation method for the random coefficient panel data model .

**Keywords**

Panel data model Random coefficient model Two-steps Aitken estimator Fixed effect estimator Group mean estimator Mean of means absolute error

## 1- Introduction

When analyzing a set of variables, many researchers are interested in arriving at the nature of the relationship between these variables. Regression analysis is one of the most important topics that studies the form of the relationship between one variable called the dependent variable and one or more variables called the independent variables, by finding a mathematical equation linking these variables, and this equation depends primarily on the values taken by the dependent variable. When the values are quantitative, most types of regression models can be used, but if the values are descriptive, there are regression models specific to this type of data. The method of collecting data varies from one phenomenon to another, which highlights different types of data. Studying a specific phenomenon for a period of time requires collecting time series data, and studying a specific phenomenon for several groups or sectors that differ among themselves requires collecting cross-sectional data, and so on according to the type of phenomenon to be studied . The emergence of this difference in the quality of the data leads to a difference in the mathematical model that represents it, if it exists, such as a linear regression model, or one of the experimental design models, or a time series model, and other models. Here it is worth noting the problems that this data suffers from. Cross-section data is characterized by its violation of the assumption of homogeneity of variances in most cases, and applying the regression analysis method to this type of data requires the use of the weighted least squares (WLS) method to estimate the model parameters . As for time series data, it gives the impression of autocorrelation since this series may be unstable, so the generalized least squares (GLS) method is used to estimate the model parameters. In both types of data above, random errors are the main reason for violating the assumptions of the regression model. With the complexity of economic and social problems and the enormous volume of statistical data expressed, statistical tools have developed thanks to informatics, which has facilitated many mathematical calculations. Perhaps the cross-sectional time series, which is called double panel data, is one of the issues for which study tools have developed and still are. Analyzing time series data separately or analyzing cross-section data separately also allows the researcher to obtain results, but he cannot be certain of their efficiency. This calls for obtaining another type of data by combining time series data and cross-section data, which results in panel data.

## 2- Panel Data[2][5]

The term panel data is Time Series and Cross Section Data, which is generally an analysis of longitudinal data. Panel data is used in many fields, such as medicine, sociology, economics, and other fields. It provides a good and improved method for studying a particular phenomenon due to better estimations than if time series data and cross-section data were used separately. One of the advantages of using paired data is that it takes into account individual differences, gives more useful and diverse data, has less correlation between variables and a large number of degrees of freedom, and is more efficient than time series that suffer from the problem of autocorrelation . Also, panel data has the ability to define and measure unobserved effects in Descriptive analysis and time series analysis. Based on the above, paired data can be defined as those data that are obtained through repeated observations to study a specific phenomenon for N cross-sections during a specific period of time. When studying a specific phenomenon , all cross-sections and explanatory variables are equal and for the same period of time we obtain balanced panel data . When studying the same phenomenon over a different period of time we obtain unbalanced panel data. The form of balanced panel data can be explained as follows :

**Table 1** : Balanced Panel data

| Observation | Time | Y | $X_1$ | ... | $X_K$ |
|---|---|---|---|---|---|
| 1 | 1 | $Y_{11}$ | $X_{111}$ | ... | $X_{K11}$ |
| | 2 | $Y_{12}$ | $X_{112}$ | ... | $X_{K12}$ |
| 2 | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | T | $Y_{1T}$ | $X_{11T}$ | ... | $X_{K1T}$ |
| | 1 | $Y_{21}$ | $X_{121}$ | ... | $X_{K21}$ |
| ⋮ | 2 | $Y_{22}$ | $X_{122}$ | ... | $X_{K22}$ |
| $N$ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | T | $Y_{2T}$ | $X_{12T}$ | ... | $X_{K2T}$ |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | 1 | $Y_{N1}$ | $X_{1N1}$ | ... | $X_{KN1}$ |
| | 2 | $Y_{N2}$ | $X_{1N2}$ | ... | $X_{KN2}$ |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | T | $Y_{NT}$ | $X_{1NT}$ | ... | $X_{KNT}$ |

### 3- Panel Data Models

By applying the method of regression analysis in panel data to estimate the model parameters we have several models, the first type of which is specialized with the fixed term that can be dealt with according to the two models. The first model is the *fixed effect model* and the second model is the *random effect model* . The above models focus on the constant limit and its change across cross-sections, with the change in limit tendencies in the models constant, and this topic has been discussed by several researchers in different countries. As for the second type, which is the subject of the research it assumes the randomness of marginal tendencies, so we have a Random Coefficient Regression model which has a significant impact in applied reality.

## 4- Random Coefficient Model (RCR)[4][8]

It is considered one of the common models that was proposed by the researcher Swamy 1970, assuming the randomness of the parameters (marginal tendencies) according to a simple and multiple regression model. In the panel data regression model, the parameters are non-random, but in this model the parameters are random under a certain assumption, so this model is sometimes called the random stochastic model . Assuming that there are observations of $N$ cross-sectional units and a time series $T$ the model gives the following:

$$Y_{it} = \sum_{K=1}^{K} \beta_{Kit} X_{Kit} + u_{it} \quad (1)$$

$$= \acute{\beta}_{it} X_{it} + u_{it} \quad , \quad i = 1, \dots, N \quad , \quad t = 1, \dots, T$$

By stacking equation above over time , we can rewrite(1) as follows :

$$Y_i = X_i \beta_i + u_i \quad (2)$$

where :

$$Y_i = (y_{i1}, \dots, y_{iT})' \; , \; X_i = (x_{i1}, \dots, x_{iT})' \; , \beta_i = (\beta_1, \dots, \beta_N)' \, , u_i = (u_{i1}, \dots, u_{iT})'$$

Under the following assumptions (Swamy 1970) :

**Assumption 1**: $E(u_i) = 0 \; , \; E(\, u_i \acute{u}_i) = \sigma_{ii} I_T$

**Assumption 2**: $\beta_i$ is distributed $iid$ with mean $\bar{\beta}$ and variance – covariance matrix $\Delta$.

From assumption (2) we can write :

$$\beta_i = \bar{\beta} + \delta_i \quad (3)$$

where $\bar{\beta}$ is a K×1 vector of constant that we want to estimate , $\delta_i$ is a K×1 vector of $iid$ it has following assumption :

**Assumption 3**: $E(\delta_i) = 0 \quad , \quad E\left(\delta_i \acute{\delta}_j\right) = Cov(\beta_i, \acute{\beta}_j) = \begin{cases} \Delta & , if\ i = j \\ 0 & , if i \neq j \end{cases}$

$$E\left(X_{it} \acute{\delta}_j\right) = 0 \quad , \quad E\left(\delta_i \acute{u}_j\right) = 0$$

Substituting the equation( 3) into (2) leads to:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_N \end{bmatrix} \bar{\beta} + \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & \vdots \\ \cdot & \cdot & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & X_N \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \vdots \\ \delta_N \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_N \end{bmatrix} \tag{4}$$

Which can be written as follows :

$$Y = X\bar{\beta} + \mathrm{Z}\delta + u \tag{5}$$

Let $\varepsilon = Z\delta + u$ we get :

$$Y = X\bar{\beta} + \varepsilon \tag{6}$$

where :

$Y_{NT\times 1} = (Y_1, \ldots, Y_N\acute{)}$ , $X_{NT\times K} = (X_1, \ldots, X_N\acute{)}$ , $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N\acute{)}$ , $Z_{NT\times NT} = diag(X_i)$ , $u = (u_1, \ldots, u_N\acute{)}$ , $\delta = (\delta_1, \ldots, \delta_N\acute{)}$

Then , the covariance matrix for the composite error is given by :

$$E(\varepsilon\acute{\varepsilon}) = \Omega = ZD\acute{Z} + \Sigma$$

$$D = E\left(\delta\acute{\delta}\right) = diag(\Delta, \ldots, \Delta) \tag{7}$$

$$\Sigma = E(u\acute{u}) = diag(\sigma_{11}I, \ldots, \sigma_{NN}I)$$

where Ω is a diagonal matrix whose diagonal elements are :

$$\Omega = \begin{bmatrix} Z_1\Delta\acute{Z}_1 + \sigma_{11}I & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_N\Delta\acute{Z}_N + \sigma_{NN}I \end{bmatrix} \tag{8}$$

## 5- Estimators

In this section, we will present three estimation methods for the random parameter model, as well as study the properties of these estimators. The estimators are as follows:

***5-1 Two-steps Aitken estimator***

The first step for this estimator is to estimate the mean $\bar{\beta}$ using the generalized least squares (GLS) method for model (6) as follows[1][8] :

$$\hat{\bar{\beta}}_{GLS} = (\acute{X}\Omega^{-1}X)^{-1}\acute{X}\Omega^{-1}Y \tag{9}$$

we note that the above estimator includes $\Delta$ and $\sigma_{ii}I$ $(i = 1, \dots, N)$, which are unknown parameters. So in the second step estimating these parameters by unbiased and consistent estimators :

$$\hat{\sigma}_{ii} = \frac{\hat{\acute{u}}_i\hat{u}_i}{N-K} \quad , \quad i = 1, \dots, n \tag{10}$$

$$\hat{\Delta} = \frac{S_\beta}{N-1} - \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{ii}(\acute{X}_i X_i)^{-1} \tag{11}$$

where :

$\hat{u}_i = Y_i - X_i\hat{\beta}_{i(LS)}$.

$\hat{\beta}_{i(LS)}$ is the least squares estimators form model in equation (2) as follows :

$$\hat{\beta}_{i(LS)} = (\acute{X}_i X_i)^{-1}\acute{X}_i Y_i \tag{12}$$

$$S_\beta = \sum_{i=1}^{N}\hat{\beta}_{i(LS)}\hat{\beta}'_{i(LS)} - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{i(LS)}\sum_{i=1}^{N}\hat{\beta}'_{i(LS)}$$

Substituting equations (10) and (11) into (8) obtain:

$$\widehat{\Omega} = diag\left(Z_1\hat{\Delta}\acute{Z}_1 + \hat{\sigma}_{11}I, \dots, Z_N\hat{\Delta}\acute{Z}_N + \hat{\sigma}_{NN}I\right) \tag{13}$$

Thus, we obtain a two-step Aitken estimator(2SAE) by substituting equation (13) into (9) :

$$\hat{\bar{\beta}}_{2SAE} = (\acute{X}\hat{\Omega}^{-1}X)^{-1}\acute{X}\hat{\Omega}^{-1}Y \tag{14}$$

The estimator above is consistent when $\mathrm{N} \to \infty$ and $\mathrm{T} \to \infty$ and is asymptotically when $\mathrm{T} \to \infty$ .

### *5-2 Fixed Effect Estimator*

Also called within estimator, this estimator gives a consistent estimate of the mean random parameter $\bar{\beta}$ . The basic idea of this estimator is to take the deviations of all observations of the mean time series[6].

Let $\bar{Y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ , $\bar{X}_{i.} = \frac{1}{T}\sum_{t=1}^{T} X_{it}$ and by using $(Y_{it} - \bar{Y}_{i.})$ on $(X_{it} - \bar{X}_{i.})$ then model (2) and equation (3) leads to :

$$(Y_{it} - \bar{Y}_{i.})' = (X_{it} - \bar{X}_{i.})'\bar{\beta} + (X_{it} - \bar{X}_{i.})'\delta_i + (u_{it} - \bar{u}_{i.}) \qquad (15)$$

where $\bar{u}_{i.} = \frac{1}{T}\sum_{t=1}^{T} u_{it}$

equation (15) is called *transformed model* , the fixed effect estimator of $\bar{\beta}$ is given by:

$$\hat{\bar{\beta}}_{FE} = [\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_{i.})(X_{it} - \bar{X}_{i.})']^{-1}.[\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_{i.})(Y_{it} - \bar{Y}_{i.})']^{-1} \qquad (16)$$

### 5-3 *Group Mean Estimator* [7][9]

This estimator was suggested by Pesaran & Simth(1995) for estimation of dynamic random coefficient models. This estimator gives a consistent estimate of $\bar{\beta}$ that can be obtained under general assumptions concerning and the regressors. The GME is defined as the simple average of ordinary least square OLS estimators of $\hat{\beta}_i$:

$$\hat{\bar{\beta}}_{GME} = N^{-1}\sum_{i=1}^{N}\hat{\beta}_i \qquad (17)$$

where $\hat{\beta}_i$ is defined in equation (12).

The estimator defined in (17) is asymptotically normally distributed and consistent so long as $\sqrt{N}/T \to 0$ , $N \to \infty$ *and* $T \to \infty$ [4]. Also the covariance matrix of MGE can be defined as :

$$Cov\left(\hat{\hat{\beta}}_{GME}\right) = N^{-1}.\hat{\Delta} \qquad (18)$$

Where $\hat{\Delta}$ is defined in(11).

**6. Simulation Study**

In this section, a Monte Carlo simulation will be carried out. The simulation experiments were based on generating data for the following model :

$$Y_{it} = X_{it}\beta_i + u_{it} \qquad , \quad \beta_i = \bar{\beta} + \delta_i \qquad , \beta = 1$$

$$= X_{it}\bar{\beta} + \varepsilon_i \qquad , \quad \varepsilon_i = X_{it}\delta_i + u_{it}$$

$X_{it}$ : is generated with a standard normal distribution with a mean zero and standard deviation equal to one .

$u_{it}$ : is generated with a standard normal distribution with a mean zero and standard deviation equal to one .

$\delta_i$ : is generated with the following distribution :

Model 1: generated $\delta_i$ with a uniform distribution $(-0.25\ ,0.25)$.

Model 2 : generated $\delta_i$ from normal distribution $(0, 0.5)$.

Model 2 : generated $\delta_i$ from gamma distribution $(1,1)$.

The experiment was repeated 500 to obtain accurate results and the values of T and N were chosen equal to 25, 50 and 100 . We used Mean of Means Absolute Error (MMAE) to compare between estimation methods , where :

$$MMAE = \frac{1}{r}\sum_{i=1}^{r} MAE(Y)$$

where $r = 500$ , $MAE(Y) = \sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|/n$

## 7. Simulation Results

In the results table below, the MMAE was calculated for the estimation methods and for all models as follows:

Table 1 : The MMAE for estimation methods.

| Model | (N= T=25) | | |
|---|---|---|---|
| | 2SAE | FE | GME |
| I | 0.0049 | 0.0087 | 0.4752 |
| II | 0.0234 | 0.7530 | 0.6342 |
| III | 0.1126 | 0.1248 | 0.3655 |
| ( N=T=50) | | | |
| I | 0.0012 | 0.0895 | 1.3759 |
| II | 0.4405 | 0.7830 | 0.7630 |
| III | 0.0188 | 0.3490 | 0.0222 |
| (N=T=100) | | | |
| I | 0.0045 | 0.0125 | 0.0977 |
| II | 0.1632 | 0.2927 | 0.3579 |
| II | 0.0180 | 0.3801 | 0.0244 |

## 8- Conclusions and Future Study

Through experiments with the simulation results in Table 1, the results showed that the best estimation method is the 2SAE method because it has the lowest MMAE for all values of (N,T) used . As a future study, other methods can be used such as the Bayesian method to estimate the random coefficient model.

## 9- References


1- Anh ,V. and Chelliah,T.(1999)." Estimated Generalized Least Squares for Random Coefficient Regression Models." Board of the Foundation of the Scandinavian Journal of Statistics.

2- Baltagi , B.H.(2013)." Econometric Analysis of Panel Data." 5th ed.,Chichester , John Wiley and Sons.

3- Beck,N. and Katz,J.N.(2006)." Random Coefficient Models for Time – Series – Cross Section Data :Mone Carlo Experiments of Finite Sample Properties." New York University, California Institute of Technology.

4- Hsiao,C. and Hasem ,M.(2004)." Random Coefficient Panel Data Models." Category 10, Empirical and Theoretical Methods.

5- Hsiao,C.(2004)."Analysis of Panel Data." 3rd ed.,Cambridge University Press, Cambridge.

6- Hsiao,C.Li,Q.Liang,Z.and , Xie,W.(2019)."Panel Data Estimation for Correlated Random Coefficients Models."

7- Raj,B.(1975)."Linear Regression with Random Coefficient :The finite Sample and Convergence Properties." Journal of the American Statistical Association ,70:349,127-137.

8- Swamy, P. (1970)."Efficient inference in Random Coefficient Regression Models." Springer-Verlag , New York.

9- Terry , E. Dielman.(1992)." Small Samples Properties of Random Coefficient Regression Estimator :a Monte Carlo Simulation , Communications in Statistics- Simulation and Computation .