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

Using Some Methods Estimation of Hybrid Model For Unbalanced Panel Data : Simulation Study

Hassan Hopoop Razzaq[1] **Mohammed Sadiq Abdul Razzaq**[2]

College of Administration and Economics University of Baghdad College of Administration and Economics \ University of Thi- Qar \

hasanhopop@utq.edu.iq dr_aldouri@coadec.uobaghdad.edu.iq

Abstract

In this paper, we presented a hybrid coefficients model for unbalanced panel data whose coefficients characterized by being partly fixed coefficient and another is random coefficient. The Swamy estimator (SWE) , the instrumental variable method (IV) and Bayes estimator (BE) were used to estimate the parameters of the hybrid coefficient model represented by the first slope coefficient and the mean of the random slope coefficient . different samples sizes were used in simulation including small , medium and large samples . the simulation results showed that the Swamy estimator(SWE) is better than other methods using root of mean squares error to compare between estimation methods .

Keywords

Hybrid coefficients model , Unbalanced panel data , Swamy estimator , Instrumental variables , Bayes estimators.

1- Introduction

When analyzing set of variables , many researchers are interested in determining the nature of the relationship between 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 that links these variables .this equation depends mainly on the values taken by the dependent variable . when the values are quantitative , most types of regression models can be used . however , if the values are descriptive , there are regression models specific to this type of data .

the mechanism of data collection varies from one phenomenon to another , which highlights different type 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 from each other requires collecting cross –sectional data and so on depending on the type of phenomenon to be studied . The emergence of this different in the quality of data leads to a difference in the mathematical model that represents it , if it exists , such as a linear regression model, one of the experimental design models , a time series model or other models. here it is worth noting the problems that these data suffer from , cross – sectional data are characterized by violating the assumption of homogeneity of variances in most cases and applying the regression analysis method to this type of data requires using the weighted least squares (WLS) method to estimate the model parameters . as for the time series data , it gives the impression of the presence of autocorrelation in it , as this series may be not stationary ,therefore the generalized least squares (GLS) method is used to estimate the model parameters. in both type of data above , random errors are the main cause for violating the assumptions of the regression model .

With the complexity of economic and social problems and the huge volume of statistical data expressing them, statistical tools have developed thanks to information technology, which has facilitated many mathematical calculations . The cross – sectional time series which are called panel data are one of the issues whose study tools have developed and continue to develop. Analyzing time series data separately or cross – section data separately also allows the researcher to obtain results but he does not guarantee their efficiency . this requires obtaining another type of data by combining time series data and cross – section data which results in panel data . the term panel data is time series and cross – section data , which is generally an analysis of longitudinal data . panel data are used in many fields such as medicine , sociology , economics and other fields . they provide a good and improved method for studying a specific phenomenon due to obtaining better estimate than if time series data and cross – section data were used separately .

The main objective of this paper is estimate the parameters of a hybrid coefficient model with fixed and random slope coefficients and statement of the best estimation method for unbalanced panel data . in order to achieve the research objective , we will discuss the performance of various estimators across small , medium and large samples .

2- Literature Review

(Swamy 1971) studied statistical inference in random coefficient regression model and presented hybrid coefficient slopes model in the first time , estimating the model parameters in the presence of a random slope parameter . He also presented proposed estimators for estimating the variance of random parameter using the consistent and unbiased estimator . it was also proven that the estimates of the hybrid model parameters are the same as the maximum likelihood method when both the model error and the random slope error are normally distributed .

(Amemiya1978) estimated the hybrid coefficient model using the least squares method when the random slope coefficient is itself a dependent variable and contains an independent variable. He also showed that such a model arises in econometric applications when combining time series data with cross – sectional data. it also proves the equivalence of the estimator in the general model and how a special case of the general model arises .

(Raj , Srivastava and Ullah1980) presented a study that included estimating a regression with random and fixed coefficient (hybrid) was estimated using estimation methods takes the form of a modified two stage least squares estimator this estimators concluded general two stage least squares random (GTSLSR) and general two stage least squares random modified (GTSLSRM) .the estimators were fitted to united states (US) data for the period (1921-1941) using Klein’s model I , where the first slope coefficient and the mean of the random slope coefficient were estimated . the result showed that the (GTSLSRM) estimator is less efficient than the (GTSLSR) estimator , they also concluded that the proposed estimators (GTSLSR and GTSLSRM) are better than the classical (TSLS) estimator .

(Balestra and Negassi 1992) estimated the parameters of a hybrid coefficient model represented by the first slope coefficient and the mean of the random slope coefficient of foreign investment of French firms using the instrumental variables methods . the estimation methods was applied to a committee of 64 factories belonging to17 French companies during the periods (1976 – 1982) . they concluded that the study empirically confirms the findings of many economists and French multinational companies with small and medium enterprises .

(Abonazel 2018) presented a paper in which he discussed panel data models when the errors are first order serially correlated as well as with random slope coefficients . the generalized least squares (GLS) estimator was used with a proposed estimator is simple mean group (SMG) to compare the random , mixed and fixed coefficient models . using Monte Carlo study the results showed that the (SMG) estimator is more reliable than the (GLS) estimators , especially for the mixed coefficient model .

3- Materials and Methods

3-1 Panel Data Concept

Panel data or is called cross – sectional time series data , as it consists of time series data and cross – sectional data is known as a technique that combines the characteristics of both cross – sectional and time series data [9]. cross – sectional data describes the behavior of a number of observations or cross –sectional units over a single period of time , while time series data describes the behavior of a single item during a specific period of time . that is panel data is hybrid data by combining two types of data where it can combine the advantages of each . that is contains a time series for each cross- sectional data for each observation in the sample under study . it is also possible to some degree to avoid the shortcomings that exist in each[6] .

meant by panel data is cross – sectional observation such as (countries , cities , companies,…) observed over a period of time , that is merging cross – sectional with time series data at the same time [10].the observations in panel data models are of two types , when the cross – sectional observations are measured for the same time periods then they are called *balanced panel data* , if the time periods differs during one or more cross – sections then they are called *unbalanced panel data* [4] . in the case of balanced , the effect of time can be neglected and the average can be subtracted . however in the case of unbalanced and different time periods the model can be estimated in the presence of the effect of time and the differences are also neglected . the panel data structure can be explanation for N cross – sections and a time series T with K independent variables as in table (2-1) shown below[14] :

Table (1) panel data structure.

Cross – Section (Individuals)	Time	Y	X_1	X_2	...	X_K
1	1	Y_{11}	X_{111}	X_{211}	...	X_{K11}
	2	Y_{12}	X_{112}	X_{212}	...	X_{K12}
	⋮	⋮	⋮	⋮	⋮	⋮
	T	Y_{1T}	X_{11T}	X_{21T}	...	X_{K1T}
2	1	Y_{21}	X_{121}	X_{221}	...	X_{K21}
	2	Y_{22}	X_{122}	X_{222}	...	X_{K22}
	⋮	⋮	⋮	⋮	⋮	⋮
	T	Y_{2T}	X_{12T}	X_{22T}	...	X_{K2T}
⋮ N	⋮	⋮	⋮	⋮	⋮	⋮

we note from the table above that the time periods is the same for all cross – sectional observations , so is called balanced panel data meaning that the time series is expressed as $(t = 1,2, \dots, T)$ and in this case the sample size is equal to (n=NT). but the case of unbalanced panel data the time period varies from one individual section to another and the time expressed as $(t = 1,2, \dots, T_i)$ and the sample size in this case is equal to $(n^* = \sum_{i=1}^{N} T_i)$.

3-2 Hybrid Coefficients Model

In classical linear regression analysis , the regression coefficients are assumed to be fixed over the sample in the panel data , that mean they do not change from one observation to another in the cross sections [2] . this means that the parameters of the panel data regression model are non-random .it can said that this assumption is rather restrictive. if panel data are used for all relevant study variables , it is very likely that the regression coefficient will not remain the same over the sample period . in such a model , we assume that the slope coefficient parameters has a probability distribution and in this sense it is a random variable. then the model is called random coefficient model(RCM) and it is called random variable model [9]:

$$Y_{it} = \sum_{k=1}^{K} \beta_{ki} X_{kit} + \varepsilon_{it} \quad \ldots \quad (1)$$

where $i = 1,2,\ldots, N$ and $t=1,2,\ldots,T$ refers to cross sectional unit and time series period respectively, $K < N$ and $K < T$ where K represents the number of independent variables . When the $(K \times 1)$ slope coefficient vector β_i for each cross section unit is considered random vector , then the above model becomes the random in order to allow for such a possibility , a model containing fixed and random slope coefficients is needed .

the regression model with both fixed and random coefficients can be represented by [1][15]:

$$Y_{it} = X_{1it}\beta_1 + X_{2it}\beta_{2i} + u_{it} \qquad i = 1,2,\ldots, N \quad t = 1,2,\ldots,T_i \quad \ldots \quad (2)$$

However, if the relationship between the dependent variable and any explanatory variable is fixed among all individuals , then RCM is not appropriate in this case . the term "fixed" used here to mean that the values of the coefficients are equal for all individuals.

model (2) is unbalanced panel data , meaning that there are N individuals observed over varying time periods length (i=1,2, …, T_i). the model in (2) can be rewritten by stacking over time period t [2][13]:

$$Y_i = X_{1i}\beta_1 + X_{2i}\beta_{2i} + u_i \qquad i = 1,2,\ldots, N \quad \ldots \quad (3)$$

where :

Y_i : a column vector of rank (T_i×1) from observations of the dependent variable for cross sectional i .

X_{1i}: matrix of rank ($T_i \times K_1$) from observations of the explanatory variable on K_1 for cross sectional i .

X_{2i}: matrix of rank ($T_i \times K_2$) from observations of the explanatory variable on K_2 for cross sectional i . where $K_1 + K_2 = K$.

β_1 : a column vector of rank ($K_1 \times 1$) of fixed (nonrandom) regression slope coefficients for cross sectional i .

β_{2i} : a column vector of rank ($K_2 \times 1$) of random regression slope coefficients for cross sectional i .

u_i : vector of random errors of rank ($T_i \times 1$) .

the hybrid coefficients model has the following assumptions [15][16]:

(1) the sample size of each cross section is greater than $K=K_1+K_2$.

(2) X_{1i} and X_{2i} are nonstochastic , the rank of X_{1i} and X_{2i} are K_1 , K_2 respectively .

(3) u_i (i=1,2,…, N) is independently and identically distributed with :

$$E(u_i) = 0 \quad , \quad E(u_i u_j') = \begin{cases} \sigma_{ii} I_T & if \quad i = j \\ 0 & if \quad i \neq j \end{cases}$$

(4) the regression slope coefficient vectors β_{2i} are independently and identically distributed with:

$$E(\beta_{2i}) = \overline{\beta}_2 \quad , \quad E(\beta_{2i} - \overline{\beta}_2)(\beta_{2j} - \overline{\beta}_2)' = \begin{cases} \Delta & if \quad i = j \\ 0 & if \quad i \neq j \end{cases} \qquad i, j = 1,2,..,N$$

(5) The β_{2j} and u_i are independent for every i and .

(6) β_{1i} ($i = 1,2, \dots, N$) are the fixed slope coefficient vectors .

(7) β_i and β_j are independent for $i \neq j$.

(8) β_i and u_j are independent for $i \neq j$.

from assumption (4) the slope coefficient are randomly distributed with mean $\bar{\beta}_2$ and variance Δ , and β_{2i} can be rewritten as[7] :

$$\beta_{2i} = \bar{\beta}_2 + \delta_i \quad , \quad i = 1,2,\ldots,N \qquad \ldots (4)$$

where :

$\bar{\beta}_2$: vector of the common constant mean coefficient across all i of rank $K_2 \times 1$.

δ_i : vector of random effect of rank $K_2 \times 1$ satisfy the following assumption :

$$E(\delta_i) = 0 \quad , \quad E(\delta_i \, \delta_j') = \begin{cases} \Delta & if \quad i = j \\ 0 & if \quad i \neq j \end{cases} \qquad i, j = 1,2,\ldots,N$$

where Δ is the variance-covariance matrix of rank($K_2 \times K_2$) , also $E(\delta_i u_j) = 0$, $\forall$ i and j.

4- Estimation Methods

As we mentioned previously , the goal is to estimate the parameter of the hybrid coefficient model for unbalanced panel data represented by the fixed first slope β_1 coefficient and the mean of the random slope coefficient $\bar{\beta}_2$.

4-1 Swamy Estimator (SWE)

Swamy explained form the assumptions of the hybrid model as well as the random parameter assumption that the problem is how to estimate the parameters $\bar{\beta}_2, \beta_1$, Δ, σ_{ii} . by substituting assumption (4) into equation(3) , the model can be rewritten for N cross-sections as follows by using matrices[16] :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} X_{11} & 0 & \cdots & 0 \\ 0 & X_{12} & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & X_{1N} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{1N} \end{bmatrix} + \begin{bmatrix} X_{21} \\ X_{22} \\ \cdot \\ \cdot \\ \cdot \\ X_{2N} \end{bmatrix} \bar{\beta}_2 + \begin{bmatrix} X_{21} & 0 & \cdots & 0 \\ 0 & X_{22} & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & X_{2N} \end{bmatrix} \begin{bmatrix} \delta_{21} \\ \delta_{22} \\ \cdot \\ \cdot \\ \cdot \\ \delta_{2N} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_N \end{bmatrix}$$

or

$$Y = D_{X_1}\beta_1 + X_2\overline{\beta}_2 + D_{X_2}\delta_2 + u \quad \ldots (5)$$

where :

$$Y_{(n^* \times 1)} = [Y_1', Y_2', \ldots, Y_N']', \quad y_{i_{(T\times 1)}} = [y_{i1}, y_{i2}, \ldots, y_{iT}]', \quad D_{X_{1_{(n^*\times NK_1)}}} = diag(X_{1i})$$
$$\beta_{1_{(NK_1\times 1)}} = [\beta_{11}', \beta_{12}', \ldots, \beta_{1N}']', \quad X_{2_{(n^*\times K_2)}} = [X_{21}', X_{22}', \ldots, X_{2N}']', D_{X_{2_{(n^*\times NK_2)}}} = diag(X_{2i})$$
$$\delta_{(NK_2\times 1)} = [\delta_{21}', \delta_{22}', \ldots, \delta_{2N}']', \quad u_{n^*\times 1} = [u_1', u_2', .., u_N']'.$$

by rewriting the model (5) we obtain the following :

$$Y = D_{X_1}\beta_1 + X_2\overline{\beta}_2 + \varepsilon \quad \ldots (6)$$

where :

$$\varepsilon = D_{X_2}\delta_2 + u$$

where the variance – covariance matrix of the composite error in(6) under the assumptions of the hybrid model is given by[2] :

$$E(\varepsilon\varepsilon') = D_{X_2}(I_N \otimes \Delta_{\beta_2})D_{X_2}' + (\Sigma_u \otimes I_T) = H_2(\theta_2) \quad \ldots (7)$$

where :

$$\Sigma_u = diag(\sigma_{ii} I_T).$$

θ_2 **:** is vector of rank $\frac{1}{2}(K_2(K_2+1)+2N)) \times 1$ containing all unknown parameter of Δ_{β_2} and $\sigma_{ii} I_{T_i}$

can also write the covariance matrix for the composite disturbance term(ε) by:

$$H_2(\theta_2)=\begin{bmatrix} X_{21}\Delta_{\beta_2}X'_{21}+\sigma_{11}I_{T_1} & 0 & \cdots & 0 \\ 0 & X_{22}\Delta_{\beta_2}X'_{22}+\sigma_{22}I_{T_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & X_{2N}\Delta_{\beta_2}X'_{2N}+\sigma_{NN}I_{T_N} \end{bmatrix} \cdots$$

(8)

where $H_2(\theta_2)$ is matrix of rank $(n^* \times n^*)$ and zeros are all $(T_i \times T_i)$ null matrices and $\Delta_{\beta_2}$is the variance – covariance matrix of β_{2i} .

applying swamy estimator to (6) we obtain :

$$\begin{bmatrix} \hat{\beta}_{1(SWE)} \\ \hat{\beta}_{2(SWE)} \end{bmatrix} = \begin{bmatrix} D'_{X_1}H_2(\theta_2)^{-1}D_{X_1} & D'_{X_1}H_2(\theta_2)^{-1}X_2 \\ X'_2H_2(\theta_2)^{-1}D_{X_1} & X'_2H_2(\theta_2)^{-1}X_2 \end{bmatrix}^{-1} \begin{bmatrix} D'_{X_1}H_2(\theta_2)^{-1}Y \\ X'_2H_2(\theta)^{-1}Y \end{bmatrix} \quad \cdots$$

(9)

in the equation (9) the matrix $H_2(\theta_2)$ is unknown , Swamy suggests to estimate this matrix by using unbiased and consistent estimators for $\Delta_{\beta_2}$and σ_{ii} as follow:

$$\hat{\Delta}_{\beta_2} = \frac{S_{\hat{\beta}_2}}{N-1} - \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2(X'_{2i}M_{1i}X_{2i})^{-1} \quad \cdots$$

(10)

where :

$$S_{\hat{\beta}_2} = \sum_{i=1}^{N}\hat{\beta}_{2i}\hat{\beta}'_{2i} - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{2i}\sum_{i=1}^{N}\hat{\beta}'_{2i}$$

$$\hat{\beta}_{2i} = (X'_{2i}M_{1i}X_{2i})^{-1}(X'_{2i}M_{1i}Y_i) \quad \cdots$$

(11)

$$M_{1i} = I_{T_i} - X_{1i}(X'_{1i}X_{1i})^{-1}X'_{1i}$$

and

$$\hat{\sigma}_i^2 = \frac{\hat{u}'_i\hat{u}_i}{T_i-K} = \frac{Y'_iM_{1i}Y_i}{T_i-K} \quad \cdots$$

(12)

4-2 *Instrumental Variables Estimator* (*IVE*)

The hybrid coefficients model in(3) can be written conveniently as by[12]:

$$Y_i = Z_i\beta_i + u_i \qquad i = 1,2,\dots,N \qquad \dots(13)$$

where Y_i and u_i are defined in equation (3) , $Z_i = (X_{1i}, X_{2i})$ and $\beta_i = (\beta_1', \beta_{2i}')'$where β_1, β_{2i} represent stochastic and non-stochastic parameter also defined in(3).under assumption (3) we can write N individuals equations by combined as :

$$Y = Z\overline{\beta} + \varepsilon \qquad \dots(14)$$

where $Z = (Z_1', \dots, Z_N')'$, $\overline{\beta} = (\beta_1', \overline{\beta}_2')'$ and $\varepsilon = (\varepsilon_1', \dots, \varepsilon_N'), \varepsilon_i = X_{2i}\delta_i + u_i$ .the method(IVE) was suggest by Balestra and Negassi (1992) to estimate the slope coefficients of the hybrid panel data model[9] . this method also give consistent estimates of the model parameters .the (IVE) using the information on all cross – sectional can be shown as follows[5] :

$$\hat{\overline{\beta}}_{IVE} = (\sum_{i=1}^{N} Z_i'F_iZ_i)^{-1}\sum_{i=1}^{N} Z_i'F_iZ_i\hat{\beta}_{i(IVCL)} \qquad \dots(15)$$

where $\hat{\beta}_i$ is the classical (IV) estimator which uses X_{2i}as instrumental and by (OLS)

for each cross section i this give :

$$\hat{\beta}_{i(IVCL)} = (Z_i'X_{2i}(X_{2i}'X_{2i})^{-1}X_{2i}'Z_i)^{-1}Z_i'X_{2i}(X_{2i}'X_{2i})^{-1}X_{2i}'Y_i \qquad \dots(16)$$

and

$$F_i = X_{2i}(X_{2i}'(X_{2i}'\hat{\Delta}_{\beta_2}X_{2i}' + \hat{\sigma}_{ii}I_T)X_{2i})^{-1}X_{2i}' \qquad \dots(17)$$

where :

$$\hat{\sigma}_{ii} = \frac{1}{T_i} \hat{u}_i' \hat{u}_i$$

is consistent , and

$$\hat{u}_i = Y_i - Z_i \hat{\beta}_{i(IVCL)}$$
$$\hat{\Delta}_{\beta_2} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\beta}_{2i} - \hat{\beta}_{2(MG)})(\hat{\beta}_{2i} - \hat{\beta}_{2(MG)})'$$

where $\hat{\beta}_{2i}$ is the sub vector of the (IVCL) estimator in equation (16) and $\hat{\beta}_{2(MG)}$ is called mean group(MG) estimator, given by :

$$\hat{\beta}_{2(MG)} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_{2i}$$

4-3 *Empirical Bayes Estimator* (*EBE*)

Under assumption that (u_i, δ_i) are independently normal distribution , it is possible apply the (EBE) to estimate $(\bar{\beta})$ conditional on (σ_{ii}) and $(\Delta_{\hat{\beta}_2})$ is equal to[8] :

$$\hat{\bar{\beta}}_{EPE} = \left\{ \sum_{i=1}^{N} (\hat{\sigma}_i^2 (Z_i' Z_i)^{-1} + \hat{\Delta}_{\beta_2})^{-1} \right\}^{-1} \sum_{i=1}^{N} (\hat{\sigma}_i^2 (Z_i' Z_i)^{-1} + \hat{\Delta}_{\beta_2})^{-1} \hat{\beta}_i \quad \ldots$$

(18)

where :

$$\hat{\sigma}_i^2 = \frac{(Y_i - Z_i \hat{\beta}_i)'(Y_i - Z_i \hat{\beta}_i)}{T_i - K}$$

…(19)

$(\hat{\beta}_i)$ is the ordinary least squares estimators by regression Y_i on Z_i for each cross section i which equal to :

$$\hat{\beta}_i = (Z_i' Z_i)^{-1} Z_i' Y_i$$

…(20)

Let

$$(Z_i'Z_i)^{-1} = \begin{bmatrix} Z_i^{11} & Z_i^{12} \\ Z_i^{21} & Z_i^{22} \end{bmatrix}$$

then an unbiased estimator of $(\Delta_{\hat{\beta}_2})$ is[12] :

$$\hat{\Delta}_{\beta_2} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\beta}_i - N^{-1}\sum_{i=1}^{N}\hat{\beta}_i)(\hat{\beta}_i - N^{-1}\sum_{i=1}^{N}\hat{\beta}_i)' - \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2(Z_i^{22}) \quad \ldots (21)$$

we note that from formula (18) that the (EBE) is the weighted average of the ordinary least squares method of cross sections , which can be rewritten by using weights[11] :

$$\hat{\hat{\beta}}_{EPE} = \sum_{i=1}^{N} W_i \hat{\beta}_i \quad \ldots (22)$$

where :

$$W_i = \left\{\sum_{i=1}^{N}(\hat{\sigma}_i^2(Z_i'Z_i)^{-1} + \hat{\Delta}_{\beta_2})^{-1}\right\}^{-1}(\hat{\sigma}_i^2(Z_i'Z_i)^{-1} + \hat{\Delta}_{\beta_2})^{-1}$$

is a positive definite matrix satisfying $\sum_{i=1}^{N} W_i = 1$ .the variance – covariance matrix of $(\hat{\hat{\beta}}_{EPE})$ under hybrid model assumption is :

$$\mathrm{var}(\hat{\hat{\beta}}_{EPE}) = \left\{\sum_{i=1}^{N}(\hat{\sigma}_i^2(Z_i'Z_i)^{-1} + \hat{\Delta}_{\beta_2})^{-1}\right\}^{-1}$$

where $\hat{\sigma}_i^2(Z_i'Z_i)^{-1}$ is the variance of (β_i)which is defined in equation(13). the estimator (18) is called empirical Bayes estimator because obtained by replacing the σ_{ii} , $\Delta_{\beta_{2i}}$in(18) with consistent estimates (19) and (21) [10].

5- The Simulation Studies

The simulation experiments were based on generating data for the following hybrid model :

$$Y_{it} = \beta_1 X_{1it} + \beta_{2i} X_{2it} + u_{it}$$

for $i = 1,2, \dots, N$ $t = 1,2, \dots, T_i$. the explanatory variables and the random errors in the equation above were generating according to a number of specifications to provide a variety of conditions under which they can be estimated . also the slope coefficients were generated one fixed and the other random .

under the assumptions of the hybrid coefficient model (23) simulation can be performed to generate its variables as follows :

(1) The values of explanatory variables (X_{1it} , X_{2it}) were generated as independent random variables that follow a normal distribution with mean μ_X and standard deviation σ_X . allowing the values of the explanatory variables to differ for each cross – sectional unit , these values are fixed across all Monte Carlo trials if they are generated for all N cross – sectional unit **.** the value of μ_X equal to zero and the value of σ_X equal to one .

(2) The random slope coefficient β_{2i} was generated from the assumption (4) and is equal to $\beta_{2i} = \bar{\beta}_2 + \delta_i$ where $\bar{\beta}_2 = 0.4$ and $\delta_i$were generated as normal distributed with mean μ_δ equal to zero and a variance Δ_δ equal to 25 , 30 . the fixed slope coefficient is chosen equal to $\beta_1 = 6$.

(3) The error term u_{it} were generated as independent normally distributed with mean set equal to zero , and variance σ_u^2 were equal to 2,4,6 and12 . The error terms are independent of the explanatory variables , and they also allowed to differ for each cross – sectional unit .

(4) Different values of N and T were chosen to be 6 , 8 , 10 , 12 , 14 , 16 , 18 and 20 to represent small , medium and large samples of individuals in each cross section and number of cross section . where the value 6 and 8 represent small samples and the values 10 , 12 and 14 represent medium samples , while the values 16 , 18 and 20 represent large samples . note that for each N cross sections chosen above , there is a cross section observed over varying time periods length (T_i =4 , i=3) , this means that the third cross section contains a time period of four individuals to obtain unbalanced panel data .

(5) All simulation experiments were repeated 1000 items for each experiment . based on all experiments, results were obtained comparisons were made between estimation methods and the optimal methods were determined .

(6) The parameters β_1 and β_{2i} have several different values chosen which allow studying the estimators . five different values of parameters are used as given in table (2).

Table* (2) *values of fixed slope coefficient and random slope coefficient.

Model	$\boldsymbol{\beta_1}$	$\boldsymbol{\bar{\beta}_2}$	$\Delta_{\beta_{2i}}$
I	2	0.1	25
II	4	0.2	30
III	6	0.4	25

6- Simulation Results

In order to compare the estimation methods for the hybrid coefficient model with unbalanced panel data , we used to selected the best estimation methods :

6-1 Root Mean Squares Error (RMSE)

$$RMSE = \sqrt{MSE} \qquad \dots (24)$$

where :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Tables (3) into (5) included a display of the values root of mean squares error (RMSE) for the SWE , IVE and EBE estimators with all models used in the simulation .

Table* (3) *RMSE for all estimation methods when $\boldsymbol{\beta_1 = 2}$, $\boldsymbol{\beta_{2i} \sim N(0.1\ , 25)}$

$\boldsymbol{N = T}$	***Estimators***	$\boldsymbol{\sigma_u^2}$			
		2	**4**	**6**	**12**
6	SWE	1.21493	1.28213	1.39462	2.05257
	IVE	3.50986	3.58416	3.81787	4.78603
	EBE	1.44852	2.03762	2.27692	3.86391
	SWE	0.91980	1.05224	1.20685	1.69123

8	IVE	3.03879	3.23537	3.42733	4.44565
	EBE	1.08360	1.58487	1.73751	2.83146
10	SWE	0.9198	1.0522	1.2069	1.6912
	IVE	3.0388	3.2354	3.4273	4.4457
	EBE	1.0836	1.5849	1.7375	2.8315
12	SWE	0.86450	0.91677	1.03231	1.38154
	IVE	2.94612	3.25344	3.27516	4.09131
	EBE	0.86291	1.15727	1.64418	2.16873
14	SWE	0.80021	0.86228	0.90775	1.24444
	IVE	2.97232	3.12127	3.31540	3.72943
	EBE	0.90270	1.07836	1.26943	2.08532
16	SWE	0.73509	0.78343	0.81867	1.09874
	IVE	2.89326	3.05858	3.06984	3.70068
	EBE	0.85979	0.95135	1.19230	1.69813
18	SWE	0.70547	0.73921	0.81196	0.99686
	IVE	2.82055	2.78795	3.13162	3.58793
	EBE	0.76094	0.93077	0.99464	1.42765
20	SWE	0.65563	0.69523	0.73708	0.92386
	IVE	2.82368	2.80344	3.13405	3.51242
	EBE	0.69512	0.60116	0.91851	1.28666

From table (3) , the results showed that the best estimation method is the Swamy estimator (SWE) , as it has the less RMSE for all samples size and variances , but for ($N = T = 12$) and $\sigma_u^2 = 12$ where the results showed the empirical Bayes (EB) estimator the best estimator , also when ($N = T = 20$) and $\sigma_u^2 = 12$ the result showed the (EB) estimator has the less RMSE. As for the instrumental variables (IV) method , it came in last place because it has the largest RMSE for all values of N, T and σ_u^2 . we also note that the RMSE values increase with increasing variance values but decrease with increasing N , T .

Figure (1) Swamy Estimator For Model I.

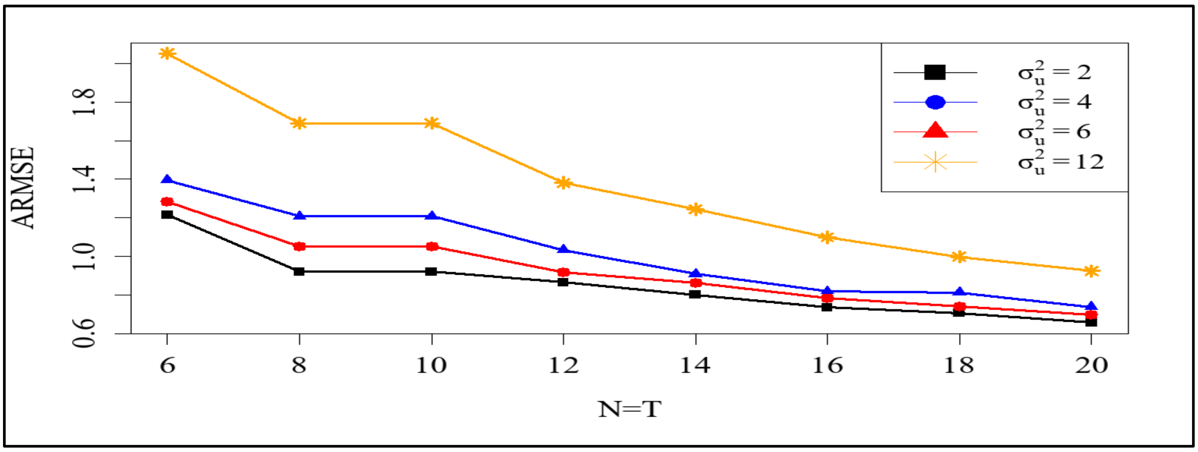


Table (4) RMSE for all estimation methods when $\boldsymbol{\beta_1 = 4 , \beta_{2i} \sim N(0.2 , 30)}$

N = T	***Estimators***	σ_u^2			
		2	**4**	**6**	**12**
6	SWE	1.30857	1.53277	1.78850	2.79839
	IVE	3.82749	4.27302	4.21578	5.70099
	EBE	1.94192	3.03579	3.75503	7.26200
8	SWE	1.26318	1.34378	1.51092	2.10259
	IVE	3.82550	3.82429	4.13073	5.16511
	EBE	1.52135	1.89442	2.34872	3.71701
10	SWE	0.98170	1.10147	1.26102	1.69120
	IVE	3.37349	3.58248	3.69171	4.50098

	EBE	1.28039	1.38678	1.80327	2.91884
12	SWE	0.87154	0.99276	1.07327	1.47086
	IVE	3.22402	3.34591	3.49354	4.28955
	EBE	0.95144	1.27277	1.55486	2.44818
14	SWE	0.87669	0.92642	0.95211	1.28041
	IVE	3.19090	3.27285	3.35988	3.86583
	EBE	1.00789	1.18516	1.55031	1.80718
16	SWE	0.81225	0.85075	0.88054	1.14665
	IVE	3.12672	3.31996	3.24459	3.87568
	EBE	0.88647	1.27370	1.29228	1.66889
18	SWE	0.77120	0.77747	0.81268	1.01026
	IVE	3.11434	3.18659	3.28913	3.82261
	EBE	0.88988	0.97250	1.06069	1.47118
20	SWE	0.71871	0.73425	0.76418	0.98493
	IVE	2.96320	3.14753	3.42030	3.68153
	EBE	0.86234	0.88153	0.96231	1.45449

From table (4) the result showed that the Swamy estimator (SWE) is best the best estimation because it has less RMSE for all value of N, T and σ_u^2. the empirical Bayes (EB) estimator came in second place as the best method , while the instrumental variables (IV) method also came in last place because it has the largest RMSE values for all values of N , T and σ_u^2. we also note that the RMSE values increase with increasing variance values but decrease with increasing N , T .

Figure (2) Swamy Estimator For Model II

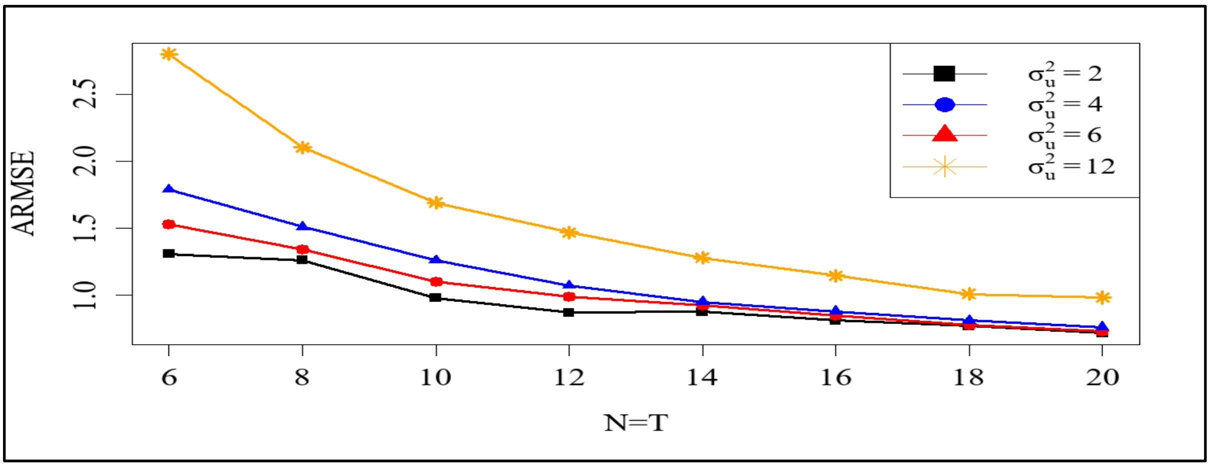


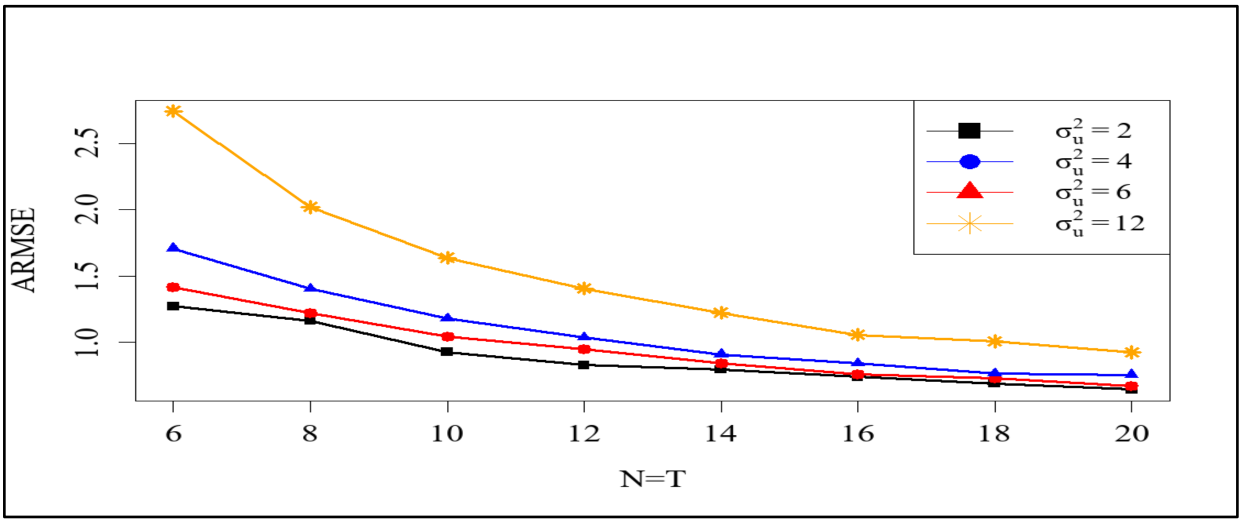
***Table* (5) *RMSE for all estimation methods when* $\boldsymbol{\beta_1 = 6}$, $\boldsymbol{\beta_{2i} \sim N(0.4\,,25)}$**

N = T	***Estimators***	$\boldsymbol{\sigma_u^2}$			
		2	**4**	**6**	**12**
6	SWE	1.27010	1.41607	1.70684	2.74300
	IVE	3.63592	3.81202	4.11241	5.71530
	EBE	1.90204	2.65878	3.86510	6.89721
8	SWE	1.16349	1.22148	1.40174	2.01987
	IVE	3.38156	3.58627	3.85501	4.82817
	EBE	1.45488	2.08334	2.56288	3.95914
10	SWE	0.92530	1.04078	1.17737	1.63652
	IVE	3.07389	3.36236	3.42432	4.47411
	EBE	1.19866	1.45164	1.79577	3.31629
12	SWE	0.82819	0.94603	1.03612	1.40480
	IVE	3.09573	3.10783	3.27628	3.96190

	EBE	0.92032	1.20765	1.50687	2.14802
14	SWE	0.79284	0.84152	0.90873	1.22088
	IVE	2.97086	2.94590	3.09523	3.69390
	EBE	0.87085	1.00350	1.17357	1.86270
16	SWE	0.73744	0.75631	0.83933	1.05565
	IVE	2.92658	3.06641	3.07699	3.61048
	EBE	0.83852	0.89448	1.03733	1.45892
18	SWE	0.68798	0.72858	0.76257	1.00712
	IVE	2.82859	3.00572	3.10046	3.46268
	EBE	0.64825	0.87151	0.97836	1.33151
20	SWE	0.64280	0.66872	0.75287	0.92367
	IVE	2.77516	2.92681	3.05407	3.43016
	EBE	0.75389	0.82743	0.97379	1.28602

From table (5), the results showed that the best estimation method is the Swamy estimator (SWE) , as it has the less RMSE for all samples size and variances , but for ($N = T = 8$) and $\sigma_u^2 = 2$ where the results showed the empirical Bayes (EB) estimator the best estimator ,. As for the instrumental variables (IV) method , it came in last place because it has the largest RMSE for all values of N, T and σ_u^2 .

Figure (3) Swamy Estimator For Model III.



7- Conclusion

In this paper , SWE , IVE and EBE estimators of hybrid coefficients model are examined when the regression parameters are random and fixed slope coefficients with unbalanced panel data . efficiency comparisons for these estimators indicate that the SWE estimator is more efficient and best than IVE and EBE estimators. we also concludes that the SWE estimator is suitable to the three models .while the IVE estimator is not suitable for three models. we also note that the RMSE values increase with increasing variance values but decrease with increasing N , T for all estimators and models .

For future studies , we recommend using other models with unbalanced panel data such as a dynamic panel data , or comparing between fixed effect model and random effect model with unbalanced panel data .

8- References

1- Abonazel , M.R.(2018).***"Different estimators for stochastic parameter panel data models with serially correlated errors".*** Journal of statistics applications & probability .J.Stat.Appl. pro.7, No.3, pp423-434 . http://dx.doi.org/10.1857/jsap/070303.

2- Abonazel ,M.R. (2018) ." ***Efficiency comparisons of different estimators for panel data models with serially correlated errors : A stochastic parameter regression approach***". international journal of system science and applied mathematics. Vol. 3,No. 2 , pp.37- 51 . doi: 10.11648/j.ijssam.20180302.14 .

3-Amemita.T,.(1987).***" A note on A random Coefficients Model."*** International Economic Review .Vol.19,No.3.

4- Algamal,Zakariya.Y. (2012).***"Selecting model in fixed and random panel data models."*** Iraq journal of statistical sciences .Iraq. No.12.pp.266-285.

5- Balestra,.P. Negassi,. S. (1992).***" A random coefficient simultaneous equation system with an application to direct foreign investment".*** Empirical economics No 17.pp 205-220. University of Geneva .Geneva.

6- Baltagi,.B. (2005).***" Econometric analysis of panel data ".*** third edition .England .

7- Beck,.N., Katz,.J.N.(2007).***" Random coefficient models for time series cross section data: Monte Carlo experiments".*** Political Analysis .No 15. pp. 182-195. Doi:10.1093/pan/mpl001.

8- Hsiao ,.C. Pesaran.M.,H. Tahmiscioglu,.A.K.(1997).***" Bayes estimation of short – run coefficients in dynamic panel data models".*** Seventh international conference on panel data , Paris .

9- Hsiao ,.C.Pesaran,.M.H,.(2004).***" Random coefficient panel data models."*** IZA DP No.1236. P.O.box 7240. Germany.

10- Hsiao,.C. (2022).***" Analysis of panel data."*** Fourth edition ,University of Southern California.

11- Hsiao,.C. Li,.Q. Liang.Z,. Xie.W,.(2019).***" Panel data estimation for correlated random coefficients models."*** Econometrics.7.7.doi:10.3390/econometrics7010007 .

12- Kalirajan,.K.P.(1990).***" On the estimation of a regression model with fixed and random coefficient."*** Journal of applied statistics, Vol 17,No.2. pp.237-244.

13- Pesaran,.M.H.(2015)." Time series and panel data econometrics." First edition Oxford university press.

14- Ratnasari ,.V. Audha .S.H,.Dani,.A.T.R,.(2023).***"Statistical modeling to analyze factors affecting the middle-income trap in Indonesia using panel data regression".*** MethodsX 11.102379. Contents lists available at Science Direct .

15-Swamy,.P.A.V.B.(1971).***"Statistical inference in random coefficient regression models."*** Ohio state university ,department of economics , Columbus,Ohio.USA.

16- Swamy,.P.A.V.B.(1970).***" Efficient inference in random coefficient regression model***". Econometrica, Vol.38,No.2,pp.311-323 .