THE WORST ANCOVA MODEL AMONG ALL POSSIBLE MODELS WITH THE SAME SET OF VARIABLES

by Suwardi Annas

THE WORST ANCOVA MODEL AMONG ALL POSSIBLE MODELS WITH THE SAME SET OF VARIABLES

I Gusti Ngurah Agung[1] and **Suwardi Annas**[2]

[1]Retired Professor of the Universitas Negeri Makassar
Sulawesi Selatan, Indonesia
Email: gusti.ng@ui.ac.id

[2]Statsitics Department, State University of Makassar
Kampus UNM Parangtambung, Jl. Daeng Tata Raya Makassar
Email: suwardi_annas@yahoo.com

ABSTRACT

*Data analysis based on two-way ANCOVA models of a numerical variable Y with categorical factors A and B, and various sets of covariates, can easily be done using a software. They are classified into two types of ANCOVA models, namely ANCOVA models with the interaction factor A*B as an independent variable, and the additive ANCOVA models without the interaction. I would consider, an additive N-way ANCOVA model is the worst model among all general linear models with the same set of variables, in a theoretical sense.*

*In addition, based on a balanced panel data with $N \times T$ individual-time or firm-time observations, various two-way fixed-effects models (TWFEMs) have been presented in the Journal of Finance (JOF). For examples, Atanassove (2013) presents several TWFEMs using over 147 thousand firm-time observations, and Vikrant (2013,p. 908) presents several TWFEMS and two three-way FEMs, namely Firm, Year, and Industry*Year Fixed Effects Models using 15,310 observation with 2,948 firms.*

In fact, a TWFEM is a special additive Two-way ANCOVA model with A is a factor of the N-firms, and B is a factor of the T-time-points, and various sets of covariates. Thence, a TWFEM is representing $N \times T$ simple linier regressions or multiple regressions with the same slopes of the covariates, and $N \times T$ difference intercepts. Furthermore, note that each of the $N \times T$ homogeneous regressions contains only a single observation, since the data has only $N \times T$ observations. So I would consider TWFEMs are the worst models. To support this statement, this paper presents an empirical results based on a data with 4×3 firm-time observations, with special notes and comments.

Furthermore, for a comparison, this paper also presents general specific equations of Heterogeneous Regressions, two-way ANCOVA models, and TWFEMs, with selected empirical outputs, using EViews. Refer to Agung (2014a, 2011) for additional examples with special notes and comments on various statistical models.

Keywords: Anacova model, Set of variable

1. INTRODUCTION

It is recognized that an ANCOVA model is a reduced a heterogeneous regressions model (HTR). which could interaction or additive *N*-way factorial ANOVA model, for *N*>1. However, I suspect, many researchers or analysts are considering there is no relationship between an ANCOVA model and a HRM. In fact, HRM might be introduced for the first time by Johson and Newman (J-N) in 1936 (quoted by Huitema, 1980).

On the other hand, many students and the researchers in the field of finance, present ANOVA models with thousands or hundred – thousand of dummy independent variables, called fixed-effects-model (FEM). For examples, Atanassove (2013) presents several TWFEMs using over 147 thousand firm-time observations, and Vikrant (2013, p. 908) presents several TWFEMS and two three-way FEMs, namely Firm, Year, and Industry*Year Fixed Effects Models using 15,310 observation with 2,948 firms.

I would consider a M-way FEM, for M > 1, as the worst ANCOVA model among all possible models with the same set of variables, in theoretical sense, since a panel data with $N \times T$ firm-time or individual-time observations would present $N \times T$ regression models with the same slope parameters. As the comparison study between the models, this paper only presents alternative simple models of the numerical problem variable *Y* by two categorical or treatment factors *A* and *B*, with a covariate *X*, such as heterogeneous regressions model (HRM), recommended ANCOVA model, not-recommended ANCOVA model, a worts ANCOVA model, and two-way fixed-effects model (TWFEM).. Note that the data of the variables *A, B, X* and *Y* can be cross-section data, experimental data, or balanced panel data (BPD). For more empirical results of a more advance models refer to Agung (2014 & 2011)

2. HETEROGENEOUS REGRESSION MODELS (HRMS)

The simplest HRM can be represented using either one of the following equation specifications (ES) in EVies. Refer to Agung (2011) for alternative ESs.

$$Y\ X*@Expand(A,B)\ \ @Expand(A,B) \tag{1a}$$

$$Y\ X\ X*@Expand(A,B,@\ Dropfirst)\ \ @Expand(A,B) \tag{1b}$$

where the function *@Dropfirst* indicates that the first cell of *(A=i,B=j),* say (1,1) is selected as the referent cell/group. Other functions can be used are *@Droplast* or *@Drop(i,j).*

On the other hand, if the joint effects the independent variables *X, A,* and *B* will be tested, it is suggested to apply one of the following alternative ESs, since the outputs directly present relevant test statistics., such as the *F*-statistic for the OLS regression, and *LR*-statistic for the binary choice models (Probit, Logit, or Extreme Value) specific for a zero-one dependent variable *Y.*

$$Y \ C \ \ X{*}@Expand(A,B) \ @Expand(A,B,@\ Dropfirst) \qquad (2a)$$

$$Y \ C \ X \ \ X{*}@Expand(A,B,@\ Dropfirst) \ \ @Expand(A,B,@@\ Dropfirst) \qquad (2b)$$

Note that the four ESs above in fact are representing exactly the same regression, with different form of functions, which can be easily obtained and written based on the outputs – see the Appendixes.

3. RECOMMENDED ANCOVA MODELS

3.1 The Simplest ANCOVA Models

The simplest two-way ANCOVA models can be represented using either one of the following equation specifications.

$$Y \ X \ \ @Expand(A,B) \qquad (3a)$$

$$Y \ \ C \ \ \ X \ @Expand(A,B,@\ ropfirst) \qquad (3b)$$

3.2 Modified Two-Way ANCOVA Models

Corresponding to the simplest ANCOVA models (3a) and (3b), the modified two-way ANCOVA models can be represented using either one of the following equation specifications.

$$G(Y) \ \ F(X) \ \ @Expand(A,B) \qquad (4a)$$

$$G(Y) \ \ C\ F(X) \ @Expand(A,B,@\ ropfirst) \qquad (4b)$$

where *F(X)* and *F(Y),* respectively, can be any functions of variable *X* and *Y,* without parameter. Refer to Agung (2014, 2011, & 2006) for alternative functions of *F(X)* and *F(Y)*

3.3 Advanced Two-Way ANCOVA Models

Corresponding to the simplest ANCOVA models (3a) and (3b), advanced two-way ANCOVA models with *K*-covariates can be represented using either one of the following equation specifications.

$$Y\ X1\ X2\ \ldots\ XK\ \ @Expand(A,B) \tag{5a}$$

$$Y\ C\ X1\ X2\ \ldots\ XK\ \ @Expand(A,B,@\ Dropfirst) \tag{5b}$$

where *Xk* can be a main variable or an interaction variable. If $Xk = X^k = X\text{^}k$ for all k=1,…,*K,* then we have a *K*-th degree polynomial ANCOVA model.

Note that having a large number of covariates, the ANCOVA models in (5) would be inappropriate models, in a theoretical sense.

4. NOT-RECOMMENDED ANCOVA MODELS

Here, I would present only the simplest ANCOVA model, which is a not-recommended model, which could be the worst ANOVA model, with the following alternative equation specifications.

$$Y\ \ X\ \ @Expand(A)\ \ @Expand(B,@Dropfisrt) \tag{6a}$$

$$Y\ \ C\ \ X\ \ @Expand(A,@\ Dropfirst)\ \ @Expand(B,@Dropfisrt) \tag{6b}$$

These models represent a set of $I \times J$ simple linear regressions (SLR), with a special pattern set of intercepts, which are known as *additive ANCOVA models.* Note that compare to previous ANCOVA models in (1) up to (5) with $I \times J$ intercept parameters, which is equal to the number of cells generated by the two factors *A* and *B*, these models only have $(I + J - 1)$ intercept parameters, which is much less than $I \times J$ for large levels of the two factors. For instance, for *I*=6, and *J*=5 these models have only 10 intercept parameters, and the recommended models have 30 intercept parameters.

So that these models have special limitations or specific pattern of intercept parameters, which are not considered and mentioned by most researchers.

As an illustration, let us see the ANCOVA models in (1a), and (6a), for *I*=2, and *J*=3. Based on the statistical results of these models, the following general equations can easily be obtained or written.

$$Y = C(1)*X + C(2)*(A=1)*(B=1) + C(3)*(A=1)*(B=2) + C(4)*(A=1)*(B=3) + C(5)*(A=2)*(B=1) + C(6)*(A=2)*(B=2) + C(7)*(A=2)*(B=3) + \varepsilon \quad (1a)$$

$$Y = C(1)*X + C(2)*(A=1) + C(3)*(A=2) + C(4)*(B=2) + C(5)*(B=3) + \varepsilon \quad (6a)$$

where (*A=i*) indicates a dummy variable or a zero-one indicator of the *i*-th level of factor *A*, and (*B=j*) indicates a dummy variable or a zero-one indicator of the *j*-th level of factor *B*.

Table-1 *Intercept parameters of the ANCOVA models in (1a) and (6a), for I=2 and J=3*

	A=1	***A=2***	***A(1-2)***
B=1	C(2)	C(5)	C(2)-C(5)
B=2	C(3)	C(6)	C(3)-C(6)
B=3	C(4)	C(7)	C(4)-C(7)
B(2-1)	C(3)-C(2)	C(6)-C(5)	***DID-1***
B(3-1)	C(4)-C(2)	C(7)-C(5)	***DID-2***

	A=1	***A=2***	***A(1-2)***
B=1	C(2)	C(3)	C(2)-C(3)
B=2	C(2)+C(4)	C(3)+C(4)	C(2)-C(3)
B=3	C(2)+C(5)	C(3)+C(5)	C(2)-C(3)
B(2-1)	C(4)	C(4)	***0***
B(3-1)	C(5)	C(5)	***0***

(b). Intercept parameters of ANCOVA Model in (6a)

Then based on these two equations, we can develop the intercept parameters of each of the models, as presented in Table-1. Based on this table the following notes and comments are presented.

(1). Each of the six cells of the Table -1(a), contains only an intercept parameter of the ANCOVA model (1a). For instance, C(2) as the coefficient of the independent dummy variable *(A=1)*(B=1)* should be in the cell *(A=1,B=1*). This table also present two statistics *Difference-In-Differences (DID)*, namely

DID-1= -C(2)+C(3)+C(5)-C(6) & DID-2=-C(2)+C(4)+C(5)-C(7)

If DID-1 ≠ 0 or DID-2 ≠ 0, then we can say that the effect of factor *A* on *Y* depends of factor *B*, or the effect of factor *B* on *Y* depends on factor *A*, which is more relevant based on a theoretical concept .

(2). The intercept parameters by the factors *A* and *B,* in Table-1(b) can be developed as follows:

- The intercept parameter C(2) as the coefficient of the dummy variable *(A=1*), it should be inserted in *(A=1,B=j),* for all *j*=1, 2, and 3.
- Similarly for the parameter C(3) as the coefficient of the dummy variable *(A=2),* it should be inserted in *(A=1,B=j),* for all *j*=1, 2, and 3.
- The intercept parameter C(4) as the coefficient of the dummy variable *(B=2*), it should be added in *(A=i,B=2),* for all *i*=1, and 2.
- Similarly for the parameter C(5) as the coefficient of the dummy variable *(B=3*), it should be added in *(A=i,B=3),* for all *i*=1, and 2.

(3). Note that Table-1(b) shows the following patterns.

- The intercept differences *A(1-2)*, say C(2)-C(3), are constant for all levels of factor *B.*
- The intercept differences *B(2-1)*, say C(4), are constant for all level of the factor *A.*
- The intercept differences *B(3-1)*, say C(5), are constant for all level of the factor *A.*
- Then we have DID-1 = DID-2 = 0.

Example-1 Application of a HRM in (1b)

Appendix-1 presents the statistical results of a HRM in (1b), based on a specific selected set of variables in Data_Faad.wf1, with its estimation command, estimation function, and a redundant variables test. Based on these results the following findings and notes area presented.

(1). The LS regression function clearly shows that the six simple linear regressions (SLRs) have different slopes, with a minimum of 2.915967 – 4.000359 = -1.084392 in (*A*=2,*B*=3), and a maximum of 3.0260033 in (*A*=2,*B*=1). In addition, note the following two specific tests.

- In the cell (A=1,B=1), X has a significant positive linear effect on Y, based on the t-statistic of t_0 = 2.509530 with df = 288 and a p-value = 0.0126/2 = 0.0063
- At the 5% level of significance, the covariate X has a significant different linear effects on Y between the two cells (A=1,B=1) and (A=2,B=3), based on the t-statistic of t_0 = -2.028769 with df = 288 and a p-value 0.0434.
- Based on these findings, I would say that an ANCOVA model is not a valid model.

(2). On the other hand, based on the *redundant variables test*, at the 10% level of significance, it is obtained a conclusion that the six slope parameters have an insignificant differences, based on the F-statistic of F_0 = 2.565710 with df = (5,288) with a p-value = 0.1697. Then, based on this conclusion, an ANCOVA model would be an acceptable model, in a statistical sense.

Example-2: Empirical Results of an ANCOVA Model

As a reduced model of the HRT in Example-1, Appendix-2 presents the statistical results of a LS regression based on the equation specification (3a), with its estimation command and estimation function, based on Data_Faad.wf1. Based on these statistical results, the following findings and notes are presented.

(1). The general equation of the ANCOVA model with six intercept parameters, and its estimation function can be easily written, using the output on the right side.

(2). The output on the left hand side presents six simple linear regressions (SLRs) of Y on X, with the same slope of $\hat{C}(1) = 1.808184$, with the first SLR has an intercept $\hat{C}(2) = 3.362735$.

(3). The covariate X has a significant positive linear effect on Y, based on the t-statistic of t_0 = 4.615031 with df = 293 and a p-value = 0.0000/2.

(4). All hypotheses on the parameter means differences, including the DID, adjusted for the covariate X, can be easily written using the parameters C(2) up to C(7), which are in fact the intercept parameters, and tested using the Wald test. For instance, a hypothesis on DID, with the null hypothesis

$$H_0: -C(2)+C(3)+C(5)-C(6) = -C(2)+C(4)+C(5)-C(7) = 0$$

is accepted based on the F-statistic of F_0 = 0.738725 with df = (2,293) and p-value = 0.4787 or the Chi-square statistic of χ_0^2 = 1.477250 with df = 2 and p-value = 0.4778. Based on this conclusion, then the interaction ANCOVA model can be reduced to an additive ANCOVA model, as presented in the following example.

Example-3 Empirical Results of a Not-recommended ANCOVA Model

Appendix-3 presents the statistical results of a LS regression based on the equation specification (6a), with its estimation command and estimation function. Based on these statistical results, the following findings and notes are presented.

(1). The general equation of the ANCOVA model with only four dummy variable and its estimation function can be easily written, using the output on the right side.

(2). The output on the left hand side also presents six simple linear regressions (SLRs) of *Y* on *X*, with the same slope of $\hat{C}(1) = 1.808184$, and special pattern of intercepts. For instance, the first SLR in (A=1,B=1) has an intercept $\hat{C}(2) = 3.439825$, the second SLR in (A=1,B=2) has an intercept $\hat{C}(2) + \hat{C}(4) = 3.439825 + 0.047629$, and the SLR in (A=1,B=3) has an intercept $\hat{C}(2) + \hat{C}(5) = 3.439825 - 0.235469$.

(3). The linear effect of the covariate *X* on *Y* is exactly the same as in the ANCOVA model (3a), with a special pattern of intercept parameters, as presented in Table-1(b), which would never be observed in reality.

Example-4 Empirical results of the worst ANCOVA Model

To present the worst ANCOVA model based on a cross-section or experimental data a fictive data set with six observation only, is generated as presented in Appendix-4, with its statistical results based on the equation specification (6a), with its estimation command and estimation function. Based on these statistical results, the following findings and notes are presented.

(1) The results present six simple linear regression of *Y* on *X* by the two factors *A* and *B*, with the same slope of 33.151116 and special pattern of intercepts as presented in the appendix.

(2) Even though its *R*-squared = 1, it does mean that the model is best possible model, since the data only has six observations for a set of six linear regressions. So each regression line only contains a single observation.

(3) So, I would say that this type of statistical results arte the worst statistical results and they doe not have any value to be presented. In addition, note that if the factor A is a set of two firms, and the factor B is the time-points, then data would be a balanced panel data (BPD) with 2 × 3 firm-time observations. To generalize, see the following section.

5. WHAT IS A TWO-WAY FIXED-EFFECTS MODEL?

I have found that a two-way fixed-effects model in fact is a two-away additive ANCOVA model based on balanced panel data. As an illustration, Appendix 5, present a subset of the data in Spcial_BPD.wf1, which is used in Agung (2014). To generalize, the variables are presented using the symbols *X1, X2*, and *Y1*, and the appendix presents the unstructured and structured panel data, called BPD.wf1, with 224×8 firm-time observations. The main objective is present to present two alternative similar statistical results, using only two estimation settings out of several or many alternative options. So the readers can conduct the analysis using any BPD using exactly the same models with other alternative options.

Examples-5 Additive Two-Way ANCOVA and Fixed-Effects Models

Appendix-6 presents two alternative stat]istical results using the following estimation methods.

(i). Using LS Estimation Method for the Additive ANCOVA Model

The LS estimation method is used to obtain the results of an additive Two-Way ANCOVA using the following equation specification, based on the unstructured BPD in Appendix-5(a). A part of the statistical results is presented in Appendix-6(a), which shows only four dummy variables of the *F_Code* out of 224 possible dummies, and seven dummy variables of the time-variable (*T*) out of 8 dummies, since the cell/level *(F_Code=1, T=1)* is used as the referent cell.

Y1 C X1 @Expand(F_Code,@Dropfirst) @Expand(t,@Dropfirst) (7)

(ii). Using Panel LS Estimation Method for the Two-Way Fixed-Effects Model

The Panel LS estimation method is used to obtain the results of the corresponding Two-Way Fixed-Effects Model (TWFEM) based on the structured BPD in Appendix-5(b), using the steps presented Appendix-7.

Based on both outputs in Appendix-6, the following findings, notes and comments are presented.

(1). They present a set of $224 \times 8 = 1792$ homogeneous regression lines *Y1* on *X1*, with a constant slope of $\hat{C}(2) = 0.005786$, and a special pattern of the intercepts, since the models have only (1+223+7) = 231 intercept parameters, which is much less than the number of cells generated by (F_Code,T), that is 1792.

(2). Since the data only has 1792 firm-time observations, which are exactly the same as the number of the homogeneous regression lines presented by both models, then I would consider that both models are the worst models, in theoretical and statistical senses. AS a comparison, refer to the statistical results and notes presented in the Example-4.

(3). It has been well known that the main objective of ANCOVA is to study the mean differences of *Y* between the cells generated by the factors, adjusted for the covariate. However the TWFEM does not present the coefficients of the independent dummy variables. So it can be said that the TWFEM in Appendix-6(b), is worse than the additive ANCOVA model presented in Appendix-6(a). Note that the model in Appendix-6(a) presents the estimates of 231 intercept parameters, but the model in Appendix-6(b) presents only one estimate of the 231 parameters, namely, $\hat{C}(1) = 1.603739$. Based on these outputs, I would say that the TWFEM is worse than the additive two-way ANCOVA model in Appendix-6(a).

(4). On the other hand note that both outputs present the same statistical values, such as the *R*-squared up to the *F*-test statistics.

(5). Based on the output in Appendix-6(a), we can present or write the set of 1792 simple linear regressions, but we can't do it based on the output in Appendix-6(b). And it is not very clear what is the value of $\hat{C}(1) = 1.603739$ in addition to the coefficient of the independent dummy variables, as the intercepts of the regression lines. Note that the output only present an equation with the symbol *[CX=F,PER=F]*.

6. Recommended Panel Data Models

Referring to the alternative simple models presented above, where the TWFEM has been considered as the worst two-way additive ANCOVA model, then I would recommend the following alternative models, specific for the balanced panel data (BPD).

6.1 Simple Heterogeneous Regression Model with Its Possible Reduced Models

It has been well known that the basic objective of data analysis is to study diffrences between group of individuals (GOI) or/and Time-Periods (TP) or time-points (T). So based on any balanced panel data (BPD) with $N \times T$ firm-time or individual-time observation, we would have to define GOI and TP, at the first stage of data analysis. Note that the categorical *GOI* should be invariant over times, and it can be generated by two of more variables, and the Time-Period should be generated based on real critical events, such as the time period before, during and after the monetary crises, and the time period with the two terror in Bali as the cutting points.

To study the differential linear effects of X on Y by a defined GOI and TP, the simplest HRM of *Y* on *X* by *GOI* and *TP*with the following equation specification would be considered.

$$Y \;\; C \;\; X \;\; X*@Expand(GOI,TP,@Dropfirst) \;\; @Expand(GOI,TP,@Dropfirst) \tag{8}$$

If the GOI should be generated based on two or more categorical factors, then it is recommended to apply the factors, instead of using *GOI*. For instance, whenever GOI should be generated by two factors, namely *A* and *B*, then the equation specification of the HRM in (8) should be presented as follows:

$$Y \;\; C \;\; X \;\; X*@Expand(A,B,TP,@Dropfirst) \;\; @Expand(A,B,TP,@Dropfirst) \tag{9}$$

Whenever the set of independent variables in *X*@Expand(GOI,TP,@Dropfirst)* have insignificant effect on *Y*, then, in a statistical sense, the interaction ANCOVA model is an acceptable model, with the equation specification as follows:

$$Y \;\; C \;\; X \;\; @Expand(GOI,TP,@Dropfirst) \tag{10}$$

Furthermore, whenever the two-way interaction factor *GOI*TP* in (10) has an insignificant effect on *Y*, then, the additive ANCOVA model would be an acceptable model, in a statistical sense only. Corresponding to two-way additive ANCOVA models, I woukd present the following notes and comments.

(1). An additive ANCOVA model should not be presented in practice, even though the interaction *GOI*TP* has insignificant effect. It is recommended to present a conclusion : *"The data does not support the hypothesis stated that the interaction GOI*TP has an effect on Y, adjusted for the covariate X"*.

(2). However, I have found many papers present Two-Way Fixed-Effects Models in the Journal of Fince, and others, which are in fact additive two-way ANCOVA models by *Firm_Code* and *Time-Point*. Refer to the papers of Atanassove (2013), and

Vikrant (2013), with their special notes presented in the introduction,which are in fact presenting $N \times T$ homogeneous multiple regressions..

(3). On the other hand, I have found an additive ANCOVA model with eight dichotomous factors and eight numerical covariates or independent variables, presented by Li, Sun, and Lettredge (2010) in the Journal of Accounting & Economics, which can be presented using a general equation specification as follows:

$$Y \ C \ X1\ X2 \ldots \ X8 \ \ DV1\ DV2 \ldots \ DV8 \qquad (11)$$

where *X1* up to *X8* are the eight selected numerical independent variables, and *DV1* up to *DV8* are the dummy variables of eight distinct dichotomous variables *V1* up to *V8*. Note that the eight dichotomous variables *V1* up to *V8* would generate $2^8 = 256$ cells or groups of individuals or firms. So the model is representing 256 multiple homogeneous regression of *Y* on *X1 –X8,* with unrealistic pattern of intercepts. Refer to the intercept parameters of a simple additive ANCOVA model in Table-1.

6.2 Heterogeneous Classical Growth Models (HCGM)

Classical growth models, namely exponential and geometric growth models, have been extensively applied in Banking and Finance, based on time series data. Here, based on BPD, I propose the simplest heterogeneous classical growth model (HCGM), as follows:

6.2.1 The Simplest HCGM by Individuals

The simplest HCGM of a positive variable Y by Individuals (I_Code) can be represented using the following equation specification. Note that log(Y) = ln(Y) in EViews.

log(Y) C t t@Expand(I_Code,@Dropfirst) @Expand(I_Code,@Dropfirst)* (12)

In practice, however, all individuals might have equal growth rates, such as interest rates, the following ANCOVA model would be appropriate model or the best possible model. In this case, the parameter C(2) indicates the fixed interest rate for all individuals.

log(Y) C t @Expand(I_Code,@Dropfirst) (13)

6.2.2 HCGM by Group of Individuals (GOI)

Since there is a possibility that some groups of individuals or firms have different interest rates, the we should consider the HCGM of a positive variable Y by GOI with the following equation specification.

log(Y) C t t@Expand(GOI,@Dropfirst) @Expand(GOI,@Dropfirst)* (14)

6.2.3 Piece-Wise HCGM by GOI and Time-Period (TP)

As an extension the HCGM of a positive variable *Y* by *GOI*, as presented in (14) we might have the interest rates are changing over some time-period (*TP*), such as before, during, and after monetary crises. Then we have to apply the following equation specification..

log(Y) C t t@Expand(GOI,TP,@Dropfirst) @Expand(GOI,TP,@Dropfirst)* (15)

7. FINAL NOTES AND COMMENTS

Based on the illustrative examples above, I would say that two-way fixed-effects models, and N-Way additive models, are the worst groups of models compare to all possible Heterogeneous Regressions and Interaction ANCOVA models , with the same set of numerical independent variables or covariates, such as the models of *Y* by *GOI* and *TP*, the models of *Y* by *GOI* and Time(*T*), and the models of Y by *I_Code* and *TP*.

For a lot of more advanced models with more numerical exogenous variable or covariate, including autoregressive models, Instrumental Variables Models, and seemingly causal models (SCM) or system equation models, refer to Agung (2014a, 2011, and 2009). Each of the books presents over 250 illustrative empirical statistical results.

REFERENCES

Agung, IG.N., 2014a. Pansel Data Analysis Using EViews. John Wiley & Sons Ltd. The Atrium, Southern Gate, Chichester, West Sussex, PO198SQ, United Kingdom

Agung, IG.N., 2014b. Manajemen Penyajian Analisis Data Sederhana untuk Skripsi, Tesis, dan Disertasi, yang Bermutu. PT RajaGrafindo Persada, Jakarta

Agung, IG.N., 2011. Cross Section and Experimental Data Analysis Using EViews. John Wiley & Sons, Inc., Singapore

Agung, IG.N., 2009. Time Series Data Analysis Using EViews. John Wiley & Sons. Inc., Singapore

Agung, IG.N., 2006. STATISTIKA: Penerapan Model Rerata-Sel Multivariat dan Model Ekonometrik dengan SPSS. Yayasan SAD SATRIA BHAKTI, Jakarta

Agung, IG.N., and Siswantoro, D. 2012. Heterogeneous Regressions, Fixed Effects and Random Effects Models, presented in the 4th IACSF International Accounting Conference. 22-23 November 2012, Bidakara Jotel, Jakarta – Indonesia

Atanassove, J. 2013. Do Hostile Takeover Stifle Innovation? Evidence from Antitakeover Legislation and Corporate Patenting. *The Journal of Finance*. Vol. LXVIII, No. 3. p. 1097-1131

Banbeko, I., Lemmon M., and Tserlukevich, Y., 2011, Employee Stock Options and Investment. *The Journal of Finance,* Vol. LXVI, No. 3, June 2011; 981 - 1009

Baltagi, B. H., 2009a, Econometric Analysis of Panel Data. John Wiley & Sons, ltd. England

Baltagi, B. H., 2009b, A Companion to Econometric Analysis of Panel Data. John Wiley & Sons, ltd. England

Bansal, P., 2005. Evolving Sustainably: A Longitudinal Study of Corporate Sustainable Development, *Strategic Management Journal,* 26: 197-218

Brooks, C. 2008. Introduction Econometric for Finance. Cambridge University Press

Ferris, K.R, and Wallace, J.S. 2009. IRC Section 162(m) and the law of unintended consequences. *Advance in Accounting, incorporating Advances in International Accounting* 25 (2009) 147-155

Filip, A. and Raffourmier, B. (2010) . "The value relevance of earning in a transition economy: The case Romania" *The International Journal of Accounting* 45 (2010) 77-103

Graybill, Franklin A. 1976. Theory and Application of the Linear Model.Duxbury Press, Belmont, California

Gujarati, D. N. 2003. Basic Econometric. McGraw-Hill, Boston.

Hair, Joseph F. JR, et.al. 2006. Multivariate Analysis. Prentice-Hall International, Inc.

Huitema, B. E. 1980. The Analysis of Covariance and Alternatives. John Wiley & Sons, New York

Lapin, L. I. 1973. Statistics for Modern Business Decisions. Harcourt Brace Jovanovich, Inc.

Li, C., Sun, L., and Ettredge, M. 2010. Financial Executive Qualifications, Financial Executive Turnover, and Adverse SOX 404 Opinions. *Journal of Accounting & Economics*

Neter, J. and Wasserman, W. 1974.. Applied Statistical Models. Richard D. Irwin, Inc. Homewood, Illinois 60430

Vikrant Vig. 2013. Acces to Collateral and Corporate Debt Structure: Evidence from a Natural Experiment. *The Journal of Finance*. Vol. LXVIII, No. 3.p. 881- 928

Wooldridge, J. M., 2002.Econometric Analysis of Cross Section and Panel Data.The MIT Press Cambridge, Massachusetts.

Appendix-1

Statistical results of a Heterogeneous Regression Model in (1a)

Dependent Variable: Y
Method: Least Squares
Date: 09/01/15 Time: 14:05
Sample: 1 300
Included observations: 300

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	2.915967	1.161957	2.509530	0.0126
X*(X*(A=1),B=2)	-1.990821	1.704891	-1.167712	0.2439
X*(X*(A=1),B=3)	-1.257814	1.332069	-0.944256	0.3458
X*(X*(A=2),B=1)	0.110066	1.402234	0.078493	0.9375
X*(X*(A=2),B=2)	-2.006040	1.603075	-1.251370	0.2118
X*(X*(A=2),B=3)	-4.000359	1.971815	-2.028769	0.0434
A=1,B=1	2.171158	1.256222	1.728323	0.0850
A=1,B=2	4.736966	1.749705	2.707294	0.0072
A=1,B=3	3.528864	1.150828	3.066369	0.0024
A=2,B=1	1.908386	0.812947	2.347493	0.0196
A=2,B=2	4.407016	1.560308	2.824452	0.0051
A=2,B=3	7.674188	2.707328	2.834598	0.0049

R-squared	0.328251	Mean dependent var	5.751100
Adjusted R-squared	0.302594	S.D. dependent var	0.945077
S.E. of regression	0.789241	Akaike info criterion	2.403688
Sum squared resid	179.3957	Schwarz criterion	2.551840
Log likelihood	-348.5532	Hannan-Quinn criter.	2.462979
Durbin-Watson stat	1.984762		

Estimation Command:
=========================
LS Y X X*@EXPAND(A,B,@DROPFIRST) @EXPAND(A,B)

Estimation Equation:
=========================
Y = C(1)*X + C(2)*X*(A=1 AND B=2) + C(3)*X*(A=1 AND B=3) + C(4)*X*(A=2 AND B=1) + C(5)*X*(A=2 AND B=2) + C(6)*X*(A=2 AND B=3) + C(7)*(A=1 AND B=1) + C(8)*(A=1 AND B=2) + C(9)*(A=1 AND B=3) + C(10)*(A=2 AND B=1) + C(11)*(A=2 AND B=2) + C(12)*(A=2 AND B=3)

Redundant Variables Test
Equation: UNTITLED
Specification: Y X X*@EXPAND(A,B,@DROPFIRST) @EXPAND(A,B)
Redundant Variables: X*@EXPAND(A,B,@DROPFIRST)

	Value	df	Probability
F-statistic	1.565710	(5, 288)	0.1697
Likelihood ratio	8.045873	5	0.1537

Appendix-2

Statistical results of an ANCOVA Model in (3a)

Dependent Variable: Y
Method: Least Squares
Date: 08/20/15 Time: 18:48
Sample: 1 300
Included observations: 300

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	1.808184	0.391803	4.615031	0.0000
A=1,B=1	3.362735	0.440156	7.639869	0.0000
A=1,B=2	3.501065	0.559719	6.255038	0.0000
A=1,B=3	3.264809	0.697009	4.684024	0.0000
A=2,B=1	3.157994	0.417076	7.571747	0.0000
A=2,B=2	3.140341	0.560808	5.599676	0.0000
A=2,B=3	2.765207	0.680012	4.066407	0.0001

R-squared	0.309992	Mean dependent var	5.751100
Adjusted R-squared	0.295862	S.D. dependent var	0.945077
S.E. of regression	0.793042	Akaike info criterion	2.397175
Sum squared resid	184.2721	Schwarz criterion	2.483596
Log likelihood	-352.5762	Hannan-Quinn criter.	2.431761
Durbin-Watson stat	1.937980		

Estimation Command:
=========================
LS Y X @EXPAND(A,B)

Estimation Equation:
=========================
Y = C(1)*X + C(2)*(A=1 AND B=1) + C(3)*(A=1 AND B=2) + C(4)*(A=1 AND B=3) + C(5)*(A=2 AND B=1) + C(6)*(A=2 AND B=2) + C(7)*(A=2 AND B=3)

Substituted Coefficients:
=========================
Y = 1.80818429155*X + 3.36273510178*(A=1 AND B=1) + 3.50106526554*(A=1 AND B=2) + 3.26480876162*(A=1 AND B=3) + 3.15799443182*(A=2 AND B=1) + 3.14034129828*(A=2 AND B=2) + 2.7652072394*(A=2 AND B=3)

Appendix-3

Statistical results of a Not-recommended ANCOVA Model in (6a)

Dependent Variable: Y
Method: Least Squares
Date: 08/24/15 Time: 20:25
Sample: 1 300
Included observations: 300

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	1.814410	0.387780	4.678962	0.0000
A=1	3.439802	0.424335	8.106336	0.0000
A=2	3.087552	0.411926	7.495397	0.0000
B=2	0.047629	0.177799	0.267879	0.7890
B=3	-0.235469	0.291083	-0.808943	0.4192

R-squared	0.306513	Mean dependent var	5.751100
Adjusted R-squared	0.297110	S.D. dependent var	0.945077
S.E. of regression	0.792339	Akaike info criterion	2.388870
Sum squared resid	185.2011	Schwarz criterion	2.450600
Log likelihood	-353.3306	Hannan-Quinn criter.	2.413575
Durbin-Watson stat	1.927381		

```
Estimation Command:
=========================
LS  Y X @EXPAND(A)  @EXPAND(B,@DROPFIRST)

Estimation Equation:
=========================
Y = C(1)*X + C(2)*(A=1) + C(3)*(A=2) + C(4)*(B=2) + C(5)*(B=3)

Substituted Coefficients:
=========================
Y = 1.81440998889*X + 3.43980153608*(A=1) + 3.08755228118*(A=2) +
0.0476286713433*(B=2) - 0.235469351523*(B=3)
```

Appendix-4

The worst statistical results of a Fictive_Data.wf1 based on ANCOVA Model in (6a)

Dependent Variable: Y
Method: Least Squares
Date: 09/02/15 Time: 14:15
Sample: 1 6
Included observations: 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	33.15116	0.020140	1646.026	0.0004
A=1	-31.79936	0.022129	-1436.978	0.0004
A=2	-32.78529	0.023807	-1377.125	0.0005
B=2	-0.991802	0.001184	-837.3934	0.0008
B=3	-1.654070	0.001429	-1157.614	0.0005

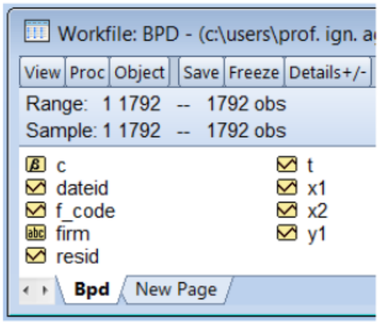
R-squared	1.000000	Mean dependent var	5.778333
Adjusted R-squared	1.000000	S.D. dependent var	1.146271
S.E. of regression	0.000762	Akaike info criterion	-11.64505
Sum squared resid	5.81E-07	Schwarz criterion	-11.81858
Log likelihood	39.93515	Hannan-Quinn criter.	-12.33972
Durbin-Watson stat	2.465116		

```
Estimation Command:
=========================
LS  Y X @EXPAND(A) @EXPAND(B,@DROPFIRST)

Estimation Equation:
=========================
Y = C(1)*X + C(2)*(A=1) + C(3)*(A=2) + C(4)*(B=2) + C(5)*(B=3)

Substituted Coefficients:
=========================
Y = 33.1511627907*X - 31.7993604651*(A=1) - 32.7852906976*(A=2)
- 0.99180232558*(B=2) - 1.65406976744*(B=3)
```

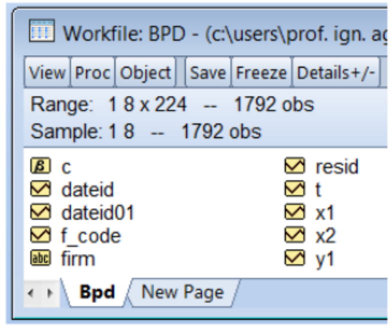
Fictive_Data				
a	***b***	***x***	***y***	***Simple Linear Regression***
1	1	1.12	5.33	$\hat{Y}$ = 33.15116*X - 31.79936
1	2	1.13	4.67	$\hat{Y}$ = 33.15116*X - 31.79936 - 0.991802
1	3	1.15	4.67	$\hat{Y}$ = 33.15116*X - 31.79936 -1.654070
2	1	1.16	5.67	$\hat{Y}$ = 33.15116*X - 32.78529
2	2	1.24	7.33	$\hat{Y}$ = 33.15116*X - 32.78529 - 0.991802
2	3	1.25	7.00	$\hat{Y}$ = 33.15116*X - 32.78529 - 1.654070

Appendix-5

Unstructured and Structured Balance Panel Data (BPD)



(a). Unstructured/Undated



(b). Structured/Dated Panel

Appendix-6

Statistical results of (a). The additive ANCOVA in (7), and (b). A Two-Way FEM of *Y1* on *X1*

Dependent Variable: Y1
Method: Least Squares
Date: 09/19/15 Time: 17:33
Sample: 1 1792
Included observations: 1792

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	***0.648507***	0.588105	1.102706	0.270300
X1	0.005786	0.012807	0.451798	0.651500
F_CODE=2	-1.00956	0.81900	-1.23268	0.21790
F_CODE=3	-0.87896	0.81902	-1.07319	0.28340
...	...	...	...	...
F_CODE=223	-0.63077	0.81900	-0.77017	0.44130
F_CODE=224	-1.00828	0.81900	-1.23111	0.21850
T=2	0.00228	0.15478	0.01471	0.98830
T=3	0.12810	0.15478	0.82763	0.40800
T=4	0.19511	0.15479	1.26043	0.20770
T=5	0.29723	0.15478	1.92041	0.05500
T=6	0.63776	0.15478	4.12056	0.00000
T=7	1.01907	0.15497	6.57610	0.00000
T=8	0.95602	0.15478	6.17682	0.00000

R-squared	0.868778	Mean dependent var	1.60368
Adjusted R-square	0.849347	S.D. dependent var	4.22008
S.E. of regression	1.637983	Akaike info criterion	3.94509
Sum squared resi	4185.461	Schwarz criterion	4.65599
Log likelihood	-3302.8	Hannan-Quinn criter	4.20757
F-statistic	44.71114	Durbin-Watson stat	0.82258
Prob(F-statistic)	0.00000		

(a)

Dependent Variable: Y1
Method: Panel Least Squares
Date: 09/19/15 Time: 17:29
Sample: 1 8
Periods included: 8
Cross-sections included: 224
Total panel (balanced) observations: 1792

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.603739	0.038694	41.44676	0.0000
X1	0.005786	0.012807	0.451798	0.6515

Effects Specification

Cross-section fixed (dummy variables)
Period fixed (dummy variables)

R-squared	0.868778	Mean dependent var	1.603677
Adjusted R-squared	0.849347	S.D. dependent var	4.220082
S.E. of regression	1.637983	Akaike info criterion	3.945090
Sum squared resid	4185.461	Schwarz criterion	4.655990
Log likelihood	-3302.801	Hannan-Quinn criter.	4.207570
F-statistic	44.71114	Durbin-Watson stat	0.460164
Prob(F-statistic)	0.000000		

Estimation Command:
=========================
LS(CX=F,PER=F) Y1 C X1

Estimation Equation:
=========================
Y1 = C(1) + C(2)*X1 + [CX=F, PER=F]

Substituted Coefficients:
=========================
Y1 = 1.60373858039 + 0.00578605726301*X1 + [CX=F, PER=F]

(b)

Appendix-7

The steps to obtain the statistical results of Two-Way FEM of *Y1* on *X1,* in Apependix-6(b), are as follow:

(1). Having the Structured BPD.wf1 on the screen, by clicking the objects: *Quick/Estimates Equation…*, then the block in Figure-1(a) shown on the screen. Then list of *Y1 C X1* can be inserted.

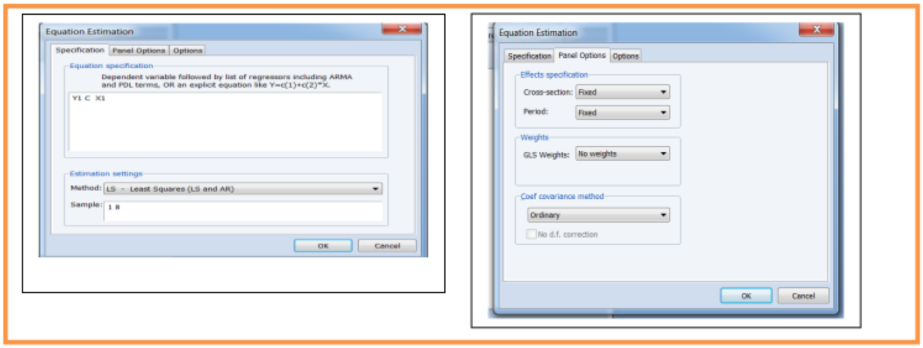


Figure-1. Equation Estimation and Panel Options for Structured Balanced Panel Data Analysis

(2). By clicking the object: *Panel Options,* then block in Figure-1(b) shown on the screen. Then by selecting the option *Fixed* for both *Cross Section* and *Period* Effects Specification, and … click *OK* directly, it is obtained the first output in Appendex-6(b). And the second output obtained by clicking the object *Representative.*

THE WORST ANCOVA MODEL AMONG ALL POSSIBLE MODELS WITH THE SAME SET OF VARIABLES

ORIGINALITY REPORT

7%	3%	8%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Agung. "ARCH Models", Time Series Data Analysis Using EViews, 12/29/2008 Publication	1%
2	sea.cc.ntpu.edu.tw Internet Source	1%
3	Kolotoğlu, Emre. "The Existence and Construction of ($K^5 \setminus e$)-Designs of Orders 27, 135, 162, and 216 : EXISTENCE AND CONSTRUCTION OF (K)-DESIGNS", Journal of Combinatorial Designs, 2012. Publication	1%
4	zombiedoc.com Internet Source	1%
5	Pratap, Rudra. "Getting Started with MATLAB", Oxford University Press Publication	1%
6	Bailey, Michael A.. "Real Stats", Oxford University Press Publication	1%

7	Deepak Kumar Panda, Rajat Kumar Bhoi. "Artificial Neural Network Prediction of Material Removal Rate in Electro Discharge Machining", Materials and Manufacturing Processes, 2005 Publication	<1%
8	Agung. "Nonlinear Least Squares Models", Time Series Data Analysis Using EViews, 12/29/2008 Publication	<1%
9	digilib.unimed.ac.id Internet Source	<1%
10	Weda, Sukardi. "Stress Shifts of English Utterances Made by Indonesian Speakers of English (ISE)", International Journal of English Linguistics, 2012. Publication	<1%
11	Timm. "Multivariate Regression Models", Springer Texts in Statistics, 2004 Publication	<1%
12	Henryk Brzeskwiniewicz. "Experiment with Split-Plot Generated by PBIB Designs", Biometrical Journal, 1994 Publication	<1%
13	www.emerald.com Internet Source	<1%
14	kups.ub.uni-koeln.de Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

THE WORST ANCOVA MODEL AMONG ALL POSSIBLE MODELS WITH THE SAME SET OF VARIABLES

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18