Choice Model between Omission of Relevant Variable and Inclusion of Irrelevant Variable in a Multicollinear Regression Model

Ijomah Maxwell Azubuike*, **Nduka Ethelbert Chinaka**

Department of Mathematics/Statistics, University of Port Harcourt, Choba, Rivers State, Nigeria

*Corresponding Author: ***Ijomah Maxwell Azubuike***, *Department of Mathematics/Statistics, University of Port Harcourt, Choba, Rivers State, Nigeria.*

Abstract: *When it comes to the actual construction of a regression model, there is little that an analyst spends more of his time wrestling with than the correct selection of variables to be included in the model. Misspecification can arise either because of omission of a variable specified by the truth, the case of the left out variable, or because of inclusion of a variable not specified by the truth, the case of the irrelevant variable. Most studies on misspecification bias has emphasized on multicollinearity as a reason for omission or inclusion of variables. This study considers both the implications of including too few ("omitted" variables) and too many ("irrelevant" variables) explanatory variables in a collinear regression model. Our result reveals that inclusion of irrelevant variables is not as severe as the consequences of omitting relevant variables in both collinear and zero correlation models.*

Keywords: *mis-specification; omitted variables; irrelevant variables; relevant variables; multicollinearity; regression model*

1. INTRODUCTION

The first assumption related to regression model is that all relevant variables should be included in the model. An economic investigation begins with the specification of the econometric model underlying the phenomenon of interest [1]. Some important questions that arise in the specification of models include what variables should be included in the model, what are the probabilistic assumptions made about the (dependent variable), (independent variable and random error term). The specification of a linear regression model consists of a formulation of the regression relationships and of statements or assumptions concerning the explanatory variables and disturbances. If any of these is violated, e.g., incorrect functional form, incorrect introduction of disturbance term in the model etc., then specification error occurs. The classical assumption that the error term is independent of the explanatory variables is violated by exclusion of a relevant variable. This error term can be seen as a collection of everything that is not accounted for by observable variables included in the model. In a classic regression equation, the estimated is little affected by omitted variables provided these are orthogonal to the remaining regressors [2]. One often faces the problem of estimating a mis-specified model in empirical research and specification problem appears to have received the most attention in empirical work by researchers such as Hoch [3], Griliches [4], and Mundlak [5]. Misspecification are the errors associated with the specification of the model, which can take many forms such as omission of relevant variable, inclusion of unnecessary variables, choosing a wrong functional form, errors of measurement etc. Omitting relevant variables from the model as a specification error has been particularly well studied relative to multiple regression analysis and the most serious consequence of this type of error is likely the biased estimates of the regression coefficients. Omitted variable bias (OVB) is one of the most common and vexing problems in ordinary least squares regression. It occurs when a variable that is correlated with both the dependent and one or more included independent variables is omitted from a regression equation [6] Even under the conditions that the estimates are not biased, the power of the relevant statistical tests is somewhat adversely affected by the omission of the relevant variables [7]. According to Clarke [8] when a model is mis specified due to omitted variable, there is always the fear of omitted variable bias, a key underlying assumption is that the danger posed by omitted variable can be ameliorated by the inclusion of control variables. The bias is a function of the omitted variables coefficient times the correlation between the included and omitted variable. Also, small amount of nonlinearity in control variables can also have a deleterious effect on the models considered [9]. The

complete regression analysis depends on the explanatory variables present in the model. In narrower sense, the specification error refers to explanatory variables. It is understood in the regression analysis that only correct and important explanatory variables appears in the model. In practice, after ensuring the correct functional form of the model, the analyst usually has a pool of explanatory variables which possibly influence the process or experiment. Generally, all such candidate variables are not used in the regression modeling but a subset of explanatory variables is chosen from this pool. In practice, the researcher is likely to have some omitted variable bias because it's impossible to control for everything that affects your dependent variable. When a model has the appropriate explanatory variables, but still fails to account for the relationship between the explanatory and response variables, then the model has functional form misspecification. If the model is mis specified then it may not be a reasonable approximation of the true data generation process. We make a functional form specification error when we choose the wrong functional form. Examples of function form misspecification include leaving out a squared variable or constraining dy/dx to be constant [10]. Specification errors can cause large forecasting errors [11], so it is of considerable importance to have means of fitting and choosing models in the presence of misspecification.

Most studies on misspecification bias has emphasized on multicollinearity as a reason for omission or inclusion of variables [12 -16]. Given the forgoing concerns, we thus need model selection techniques which will guard us against the dangers of model misspecification. Again, most researchers rarely pay attention to the other aspects of misspecification. In particular, little note is made of the consequences of irrelevant variables, or of the effects of misspecification on the variance and mean square error of the regression estimates mainly because these results are not readily available [17]. In view of the importance of these aspects of misspecification in empirical research and in order to get a clearer picture, some major results of misspecification are presented in this paper. We shall consider the classical linear regression model where all the independent variables are non-stochastic and the error terms are homoscedastic and serially independent both for ill conditioned and data free from multicollinearity. This study therefore is on model selection in the presence of misspecification. The rest of the paper is as follows: In Section 2, we consider misspecification error bias for omission of relevant variable and inclusion of irrelevant variable. Section 3 deals with the materials and methods which is followed by a numerical application in section 4 to x-ray the implications of each specification error. Section 5 concludes the work.

2. MISSPECIFICATION ERROR BIAS

Consider the model

$$y_i = \theta_1 x_{1t} + \theta_2 x_{2t} + ... + \theta_k x_{kt} + \mu_t \tag{1}$$

$$y_i = \omega_1 x_{1t} + \omega_2 x_{2t} + ... + \omega_k x_{kt} + \omega_{k+1} x_{k+1t} + \varepsilon_t \tag{2}$$

where $t = 1, 2, ..., T$

When equation (1) is the truth, but suppose for some reason a researcher decides to use equation (2), then (2) is a misspecified model because of the presence of the irrelevant variable X_{k+1}. When also equation (2) is the truth, equation (1) is a misspecified model because of the absence of variable X_{k+1}. When a relevant variable in the model is excluded, the specification error will affect the properties of OLS estimator, in the presence of such error, OLS estimators will be bias.

These two equations may be written in the matrix form as

$$Y = X\theta + \mu \tag{3}$$

$$Y = X\omega + \varepsilon \tag{4}$$

where Y is a vector of observations on the dependent variable, and X and X are matrices of independent variables in the equations (1) and (2) respectively. Without any loss of generality, the equation (3) may be rewritten as:

$$Y = \begin{bmatrix} XX_{k+1} \end{bmatrix} \begin{bmatrix} \theta \\ 0 \end{bmatrix} + \mu \tag{5}$$

where X_{k+1} is a vector of observations on the independent variable

When a relevant variable in the model is excluded, the specification error will affect the properties of OLS estimator, in the presence of such error, OLS estimators will be bias. Also when an incorrect functional form is chosen, then the model is mis specified.

2.1. Omission of a Relevant Variable

Consider the multiple linear regression model:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \mu \tag{6}$$

For some reason such as unavailability of the data on X_2, we construct a regression model without the X_2 variable

$$Y = \lambda_0 + \lambda_1 X_1 + \mu_1 \tag{7}$$

In econometrics, this issue is known as "omitting a relevant variable", if $\theta_2 \neq 0$, and this is a type of misspecification. Then, by omitting X_2, its effect becomes part of the error term in the reduced model: $\mu_1 = \mu + \theta_2 X_2$

This implies:

$$E[\mu_1 / X_1] = E[\mu + \theta_2 X_2 / X_1]$$

$$= E[\mu / X_1] + E[\theta_2 X_2 / X_1]$$

$$= E[E(\mu / X_1 X_2) / X_1] + \theta_2 E[X_2 / X_1]$$

$$= E[\mu / X_1] + E[\theta_2 X_2 / X_1]$$

$$= \theta_2 E[X_2 / X_1]$$

Hence,

$$= E[(Y / X_1] = \lambda_0 + \lambda_1 X_1 + \theta_2 E[X_2 / X_1]$$

In general, whenever we have

$$E[E(X_{1........}, X_k] \neq 0$$

Theorem 1: In the classical linear regression model, omission of a variable specified by the truth introduces bias and decreases the variance in all the least squares estimates.

Following Rao[17], Griliches [18] , The least squares estimates of the 3's in equation (1) are given by:

$$\hat{\theta} = (X'X)^{-1} X'Y \tag{6}$$

Since equation (2) is the truth we may rewrite equation (6) as:

$$\hat{\theta} = (X'X)^{-1} X'(\hat{X}\omega + \varepsilon)$$

$$= (X'X)^{-1} X'\hat{X}\omega + \varepsilon + (X'X)^{-1} X'\varepsilon \tag{7}$$

Since all the independent variables are nonstochastic, taking expectation in (7) we have

$$E(\hat{\theta}) = (X'X)^{-1} X'\hat{X}\omega \tag{8}$$

The expected value of the regression coefficient of the independent variable x1 may be written as:

The variance of the estimate vector θ

$$V(\hat{\theta}) = E\left[(\hat{\theta} - E(\hat{\theta}))(\hat{\theta} - E(\hat{\theta})) \right]$$

$$= E\left[(X'X)^{-1} X' \mathscr{E}\mathscr{E}' X (X'X)^{-1}\right]$$

$$= \sigma_\mu^2 (X'X)^{-1}$$

That is, the "omission of relevant variables" in the analysis generates inconsistency and bias in estimating the effects of variables, though a reduction in the variance of the estimator.

2.2. Inclusion of an Irrelevant Variable

Another situation that often appears is the associated with adding variables to the equation that are economically irrelevant. The researcher might be keen to avoiding the problem of excluding any relevant variables, and therefore include variables on the basis of their statistical relevance. Some of the included variables could then be irrelevant economically.

Theorem 2: In the classical linear regression model, inclusion of an irrelevant variable does not introduce bias but increases the variance in the least squares estimates

Assume that the truth be given by equation (1), and let the misspecified model, equation (2), with the irrelevant variable X_{k+1} be estimated. The least squares estimate of the misspecified model, equation (2), is given by

$$\hat{\omega} = (X'X)^{-1} X'Y \tag{10}$$

Since equation (1), which may be written without loss of generality as equation (5), is the truth, we may rewrite equation (10)

$$\hat{\omega} = (X'X)^{-1} X' \left[X \begin{bmatrix} \theta \\ 0 \end{bmatrix} + \varepsilon \right] \tag{11}$$

Therefore $E(\omega) = \begin{bmatrix} \theta \\ 0 \end{bmatrix}$ \hfill (12)

This implies: $E(\omega) = \omega$

The variance of the least squares estimates, given by equation (11) is given by

$$V(\omega) = \sigma_\varepsilon^2 (X'X)^{-1} \tag{13}$$

The variance of the least squares estimates when the true equation (1) is estimated is

$$V(\theta) = \sigma_\varepsilon^2 (X'X)^{-1}$$

i.e. $E[\varepsilon / X_1 X_2] = E[\varepsilon_1 - \omega_2 X_2 / X_1 X_2] = E[\varepsilon_1 / X_1]$

The estimates are still inconsistent and unbiased, and the only inconvenience is an increase of the residual variance and hence of the estimated standard deviation of the residual increased.

3. MATERIAL AND METHODS

The purpose of this paper is to evaluate the effect of omission of relevant variable and inclusion of irrelevant variable in a collinear and non-collinear models by using simulated incidence matrices and the COLLIN option in SAS 9.0 version PROC REG. Simulation study for investigating the effect of omission of relevant variable and inclusion of irrelevant variable in both collinear and non-collinear regression parameters consist of (x, y) vectors. The simulation study is performed as follows: Consider the standard linear regression model given as

$$Y = X\beta + \varepsilon \tag{14}$$

where: Y is an n x 1 vector of dependent variables

X is an n x k matrix of regressors

β is a k x 1 vector of parameter

e is an n x 1 vector of disturbance and is normally distributed with covariance matrix proportional to the identity matrix.

Sample data for Y, X_1, X_2 and X_3 were generated from a multivariate normal distribution (MVN) with several datasets of sample size 50, 100, 500 and 1000 with one response variable y and three predictors xi, i=1, 2, 3 were generated. The parameters of equation (14) were specified and initially fixed as $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 0.5$ and $\beta_3 = 0.5$. We introduce an irrelevant variable (X_3) by giving $\beta_3 = 50$. Thereafter, we also subjected the coefficients to $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 0.5$ and $\beta_3 = 2.5$. The levels of intercorrelation (multicollinearity) among the independent variables were determined by adjusting (u = 20, 50). By doing so in line with the objective, we obtain severe and no collinearity. To illustrate the effect of omission of relevant variable and inclusion of irrelevant variables on a collinear model, we first investigate collinearity between relevant independent variables, and then between relevant and irrelevant independent variables. We combine the findings to develop a decision matrix that researchers can utilize to determine which situation they may want to avoid, and when using it, outline strategies to detect and respond to potential multicollinearity issues. This enables a clear demonstration of the effect of inclusion of irrelevant variable to collinearity on regression and portray some conditions under which the various collinearity diagnostics are informative and some conditions where they are misleading. Secondly, we also create a model where X_1 and X_2 are correlated (relevant to each other) and introduce a third variable X_3 which is neither correlated to X_1 nor X_2 and then omit either X_1 or X_2 (Omission of relevant variable). We then observed the model selection criteria statistics like MSE, AIC, BIC, R^2.

4. RESULTS AND DISCUSSION

In this section, the result from the simulated data involving both omission of relevant variable and inclusion of irrelevant variable with three explanatory variables from multicollinear model and model free from collinearity were obtained. The criteria used for comparing evaluation of the performance of our estimators were MSE, AIC, BIC, R^2.

Case 1: Regression with inclusion of Irrelevant variable and omission of relevant variable in a collinear model

Table1. *Comparison of coefficients of the models between Inclusion of Irrelevant variable and Omission of relevant variable in a collinear model.*

N	Variable	Parameter estimate	Standard Error	t-value	Prob.	R^2	MSE	AIC	BIC	VIF
Inclusion of irrelevant variable (X_3)
50	Intercept	1.2404	0.2474	5.01	0.0000	0.999	0.8885	-7.9882	-5.3077	0
	X_1	1.0009	0.1705	4.74	0.0000					372.78
	X_2	1.0039	0.1702	6.94	0.0000					371.17
	X_3	-0.0039	0.0027	1.85	0.0703					1.0851
100	Intercept	1.2571	0.2107	5.97	0.0000	0.999	1.0922	21.5583	23.888	0
	X_1	0.9124	0.1551	5.88	0.0000					410.16
	X_2	1.0722	0.1557	6.89	0.0000					409.06
	X_3	0.0027	0.0022	1.24	0.2169					1.0778
200	Intercept	1.2021	0.1486	8.09	0.0000	0.999	1.0744	32.6459	34.808	0
	X_1	0.8749	0.1128	7.76	0.0000					460.64
	X_2	1.1126	0.1130	9.85	0.0000					460.01
	X_3	0.0019	0.0014	1.29	0.2000					1.0766
500	Intercept	1.1060	0.0916	12.08	0.0000	0.999	1.0744	32.6459	34.808	0
	X_1	0.9114	0.0693	13.15	0.0000					445.86
	X_2	1.0828	0.0693	15.62	0.0000					445.54
	X_3	0.0013	0.0009	1.26	0.1969					1.0876
1000	Intercept	1.0546	0.0643	16.39	0.0000					0
	X_1	0.9858	0.0470	20.96	0.0000					412.82

						R^2	MSE	AIC	BIC	VIF
	X_2	1.0123	0.0471	21.46	0.0000	0.999	1.0744	32.6459	34.808	412.79
	X_3	0.0003	0.0007	0.38	0.7028					1.0778
Omission of relevant variable (X₂)										
50	Intercept	1.4046	1.2862	1.09	0.2804					0
	X_1	1.2442	0.4269	2.91	0.0054	0.578	4.620	155.93	157.93	13.804
	X_3	-0.3200	0.4031	-0.79	0.4313					13.804
100	Intercept	1.4355	1.0408	1.38	0.1710					0
	X_1	1.1508	0.3447	3.34	0.0012	0.569	5.478	343.10	345.23	20.20
	X_3	-0.2874	0.3375	-0.85	0.3966					20.20
200	Intercept	1.1297	0.7469	1.51	0.1320					0
	X_1	1.0100	0.2338	4.32	0.0000	0.586	5.4308	679.81	681.81	19.38
	X_3	-0.1237	0.2282	-0.54	0.5885					19.38
500	Intercept	0.4964	0.4609	1.08	0.2820					0
	X_1	1.0304	0.1573	6.55	0.0000	0.626	5.3038	1671.41	1673.4	22.72
	X_3	-0.0800	0.1542	-0.52	0.6040					22.72
1000	Intercept	0.3306	0.3218	1.03	0.3045					0
	X_1	0.9095	0.1092	8.33	0.0000	0.640	5.2553	3321.48	3323.5	22.12
	X_3	0.0687	0.1070	0.64	0.5209					22.12

Table 1 gives an overview of the consequences of including or not including a relevant versus irrelevant collinear variable in a regression model. Naturally, not including an irrelevant variable has no detrimental consequences on regression results. When the irrelevant variable was included in the collinear model as shown in table 1, the t-value for the irrelevant variable (X_3) is insignificant as expected. The results from the exercise also shows the mean square error increased as the sample size increases. The coefficient of determination was very high (0.999) all through the samples and the VIF showed a severe collinearity. A look at the AIC and BIC indicated that under such severe collinearity, the least sample size (n = 50) outperformed others. In the case of omission of relevant variable, the coefficient of determination dropped to 0.57 but later increased slightly to 0.67 with increase in the sample. The mean square error was five times compared to omission of relevant variable and Again the smallest sample (n =50) appeared to better than others. In comparison, the omission of relevant variable shows to be more sensitive than inclusion of irrelevant variable in a collinear model.

Case 2: Regression with inclusion of irrelevant variable and omission of relevant variable but free from multicollinearity

Considering the case of inclusion of irrelevant variable and omission of relevant variable but free from multicollinearity as in table 2 below, similar effect was also demonstrated with the simulation studies when a model without multicollinearity was considered. A closer look at the table indicates that for inclusion of irrelevant variable (X_3), the coefficient of determination was very high (0.999) as in case 1. Again, the mean square error increased with increase in sample size and the sample size of 50 (n = 50) indicated a better model than with higher sample size judging from the AIC and BIC values. With omission of relevant variable (X_2), the coefficient of determination dropped slightly but later increased at n = 500 and 1000. The AIC and BIC also increased with increase in sample size. Comparing both scenario of inclusion of irrelevant variable and omission of relevant variable in a model free from collinearity, we observed that inclusion of irrelevant variable is preferred than omission of relevant variable. Furthermore, the standard error in the model for inclusion of irrelevant variable appear to be lower than omission of relevant variable.

Table2. *Regression model without collinearity*

n	Variable	Parameter estimate	Standard Error	t-value	Prob.	R^2	MSE	AIC	BIC	VIF
Inclusion of irrelevant variable (X₃)										
50	Intercept	1.2493	0.2486	5.03	0.0000					0
	X_1	0.9861	0.006	102.80	0.0000	0.999	0.8891	-7.9252	-5.2447	1.2836

Choice Model between Omission of Relevant Variable and Inclusion of Irrelevant Variable in a Multicollinear Regression Model

n	Variable	Coef	Std Err	t	p					
	X₂	1.0041	0.0057	176.92	0.0000					1.1998
	X₃	0.0050	0.0027	1.83	0.0703					1.0801
100	Intercept	1.2395	0.2145	5.78	0.0000	0.999	1.0953	22.119	24.449	0
	X₁	0.9851	0.0158	62.54	0.0000					4.6622
	X₂	1.0007	0.0164	60.86	0.0000					4.5524
	X₃	0.0025	0.0022	1.18	0.2413					1.0793
200	Intercept	1.1999	0.1453	8.26	0.0000	0.999	1.0744	32.6507	34.813	0
	X₁	0.9856	0.0058	171.44	0.0000					1.3088
	X₂	1.0021	0.0032	314.14	0.0000					1.2298
	X₃	0.0019	0.0014	1.28	0.2005					1.0757
500	Intercept	1.0862	0.0898	12.10	0.0000	0.999	1.0538	56.4175	58.482	0
	X₁	0.9921	0.0037	271.65	0.0000					1.3631
	X₂	1.0030	0.0019	528.43	0.0000					1.2807
	X₃	0.0012	0.0001	1.24	0.2146					1.0809
1000	Intercept	1.0535	0.0622	16.93	0.0000	0.999	1.0463	94.5835	96.616	0
	X₁	0.9980	0.0025	399.32	0.0000					1.3046
	X₂	1.0002	0.0013	749.62	0.0000					1.2389
	X₃	0.0003	0.0007	0.38	0.7073					1.0675
Omission of relevant variable (X₂)										
50	Intercept	1.4734	0.5341	2.76	0.2804	0.569	2.0543	74.61	74.89	0
	X₁	0.7108	0.0926	7.67	0.0000					1.0228
	X₃	0.0116	0.0199	0.58	0.5641					1.0228
100	Intercept	1.2689	0.3585	3.54	0.0006	0.565	2.0341	144.96	145.82	0
	X₁	0.7183	0.0645	11.14	0.0000					1.0208
	X₃	-0.0030	0.0132	-0.22	0.8230					1.0208
200	Intercept	1.2465	0.2652	4.70	0.0000	0.562	2.0711	294.20	295.08	0
	X₁	0.7259	0.0462	15.71	0.0000					1.0193
	X₃	0.0018	0.0092	0.20	0.8415					1.0193
500	Intercept	0.9518	0.1677	5.67	0.0000	0.595	2.036	714.26	715.02	0
	X₁	0.7680	0.0289	26.57	0.0000					1.0284

	X_3	0.0034	0.0063	0.54	0.5872					1.0284
1000	Intercept	1.0852	0.1156	9.39	0.0000	0.598	2.0497	1438.39	1439.23	0
	X_1	0.7678	0.0205	37.55	0.0000					1.0340
	X_3	0.0065	0.0044	1.50	0.1351					1.0340

5. CONCLUSION

In this paper, we studied three issues related to model misspecification bias of OLS estimators arising from errors of exclusion (of relevant variables) and inclusion (of irrelevant variables): (1) omitted variable bias; (2) possible reduction of omitted variable bias with the inclusion of some of the omitted variables; (3) selection between inclusion of irrelevant variables and omission of relevant variables using AIC and BIC criterion. In conclusion, it was found that inclusion of irrelevant variable is a safer bias than omission of relevant variable in model selection of a mis-specified linear regression model. It is clear from the result that including a collinear variable, regardless of whether it is relevant, leads to error inflation and an increase in VIFs, which makes it more difficult for the researcher to identify relevant relationships. Inclusion of irrelevant variables is not as severe as the consequences of omitting relevant variables, so the temptation is to include "everything but the kitchen sink". There is a balancing act, however, between bias and efficiency. A small amount of bias may be preferable to a great deal of inefficiency. This inclusion also assured convergence in a conditional manner. The best place to start is with good theory. Then include all the variables available that follow from this theory and then exclude variables that add least to the model and are of least theoretical importance.

REFERENCES

[1] Babatunde O.S, Ikughur A.J, Ogunmola A.O, Oguntunde P. E (2014:Effect of Omitted Variable due to Misspecification Error in Regression Analysis, International Journal of Modern Mathematical Sciences, 2014, 11(1): 49-57

[2] Cramer,J.S.,(2004): Scoring bank loans that may go wrong: a case study. Statistica Neerlandica, 58, 365-380.

[3] Hoch, L., (1962): Estimation of production function parameters combining time-series and cross-section data, Econometrica 30, 34–53.

[4] Griliches, Z. (1986): Economic data issues. In Z. Griliches & M. D. Intriligator (Eds), Handbook of econometrics, Vol. 3: 1465–1514. Amsterdam: North-Holland.

[5] Mundlak,Y.,(1961):Empirical production function free of management bias ,Journal of Farm Economics 43, 44–56

[6] Leightner, J.E., and Inoue T (2012): Solving the Omitted Variables Problem of Regression Analysis Using the Relative Vertical Position of Observations. Hindawi Publishing Corporation Advances in Decision Sciences Volume 2012.

[7] Pedhazur, E. J. (1997). Multiple regression in behavioral research: Explanation and prediction (3rd ed.). Fort Worth: Harcourt Brace.

[8] Clarke Kelvin A. (2003): Non-Parametric Model Discrimination in International Relation, Journal of Conflict resolution, 47(1), 75-80.

[9] Achen, Christopher H. (2000): "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." Annual Meetings of the Political Methodology Section of the American Political Science Association, UCLA

[10] Wooldridge, J.M. (1994): "A simple specification test for the predictive ability of transformation models," Review of Economics and Statistics 76, 59–65.

[11] White, H., (1994): Estimation, Inference and Specification Analysis. New York: Cambridge University Press.

[12] Liu, Z. and Lynk, E. L. (1999): Evidence on market structure of the deregulated US airline industry, Applied Economics, 31(9), 1083-92.

[13] Greene, C. A. (1999): On the impossibility of a stable and low GDP elasticity of money demand: the arithmetic of aggregation, replication and income growth, Applied Economics, 31(9), 1119-27.

[14] Meyer, K. E., & Sinani, E. (2009): When and where does foreign direct investment generate positive spillovers? A meta-analysis. Journal of International Business Studies, 40(7): 1075–1094

[15] Muethel, M., & Bond, M. H. (2013): National context and individual employees' trust of the out-group: The role of societal trust. Journal of International Business Studies, 44(4): 312–333.

[16] Zhao, M., Park, S. H., & Zhou, N. (2014): MNC strategy and social adaptation in emerging markets. Journal of International Business Studies, 45(7): 842–861.

[17] Rao, P., (1971): Some Notes on Misspecification in Multiple Regressions, The American Statistician, 25:5, 37-39.

[18] Griliches, Z., (1957):"Specification Bias in Estimates of Production Functions," Journal of Farm Economics, 8-20.