KIU Journal of Humanities Copyright©2018
Kampala International University ISSN: *2415-0843*; 3(4): 105–114

An Evaluation of Endogenous Model on Wage among Farming Households in Nigeria

OMOTUYOLE I. AMBALI, JOHN OLUSEGUN Y. AIHONSU,
OLUGBENGA O. OLUBANJO, KASSIM A. AKANNI
Olabisi Onabanjo University, Yewa Campus, Ayetoro, Ogun State, Nigeria

Abstract. Some variables may be endogenous in wage equation. However, this endogenous hypothesis is not usually tested in most applied studies. Therefore, this study empirically applied two stage least squares (TSLS) to examine the return to education among farming households in Nigeria using the Nigerian household living standard data of the National Bureau of Statistics (NBS) in the World Bank sponsored project. Tests of endogeneity, instrument relevance and over-identifying restriction were carried out. The test of endogeneity confirms ordinary least squares (OLS) should not be relied on as it does not give a consistent and reliable estimates. The instruments included are exogenous and not correlated with the error term, but correlated with the independent variable confirming the relevance of the instruments. The empirical result affirms education is an endogenous determinant of wage rate. One unit increase in year of education will increase the wage rate by #439.67 (Four Hundred and Thirty Nine Naira, Sixty Seven Kobo), holding other variables constant. It suggests that test for endogeneity is necessary for an unbiased and consistent estimate to be obtained. Thus ascertaining the endogenous nature of explanatory variable(s) is important to guiding the appropriate estimation method to be applied. This also affects the statistical inference and policy recommendation.

Keywords: Instrumental, Variable, Two stage least squares, Ordinary least squares, Education, Wage Rate

1. Introduction

Endogenous problem is a common econometric problem which occurs when at least one explanatory variable is correlated with the error term. The methods of instrumental variables (IV) and two stage least squares (TSLS) are used in social sciences to obtain consistent estimate due to difficulty in conducting controlled experiment, unlike in the pure sciences. An explanatory variable that is correlated with the disturbance error is called endogenous variable. This correlation is attributable to simultaneity problem, omission of relevant explanatory variable from a model, or when the covariates are subject to measurement error. If these occur, ordinary least squares

(OLS) may produce biased and inconsistent estimates (Bullock, Green and Ha, 2010). Thus the IV and TSLS are alternative methods often relied on by the social scientists.

An instrument or instrumental variable aids consistent estimates. The IV and TSLS methods are commonly used to provide consistent estimate (Angrist, Imbens, Guido and Rubin,1996). However, biased estimate is possible when there is a weak correlation between an instrument and endogenous explanatory variable (Bound, Jaeger and Baker, 1995). An instrumental variable must meet two requirements. First, it must be correlated with the endogenous explanatory variables, conditional on the other covariates. Second, an instrument must not be correlated with the error term in the explanatory equation.

The objective of this study is to ascertain whether education is an endogenous determinant of wage rate among farming households in Nigeria. This is achieved by testing the endogenous hypothesis against the alternative that education is not an endogenous determinant of wage using two-stage least squares method. The theory of instrumental variable was first derived by Philip Wright in his 1928 book entitled “The Tariff on Animal and Vegetable Oils”as cited by Stock and Trebbi (2003). Following this, IV has been applied in different contexts. Heckman (1997) equally documents the usefulness of IV in developmental programme. Leigh and Schembri (2004) noted that some variables may be correlated with health and smoking while estimating the effect of smoking on general health. Wage can be estimated as a function of education, experience, age, gender, etc. yet variables like experience influence wage, and could plausibly influence years of formal schooling. A regression of wage against education may therefore suffer from omitted variables bias. Therefore, this study extends the existing knowledge on the determinant of wage using farming household data in developing countries.

The credibility of the estimates of IV or TSLS hinges on the selection of suitable instruments. Instruments are easily created in the natural experiment but somewhat difficult to create in social sciences. For examples, Miguel, Satyanath, and Sergenti (2004) used weather shocks to identify the effect of changes in economic growth on civil conflict. On the other hands, Angrist and Krueger (2001) and Angrist and Krueger (1991) used quarter of birth as instrumental variable for ability while Angrist and Krueger (1992) noted that parenting is a key omitted variable in a wage equation. Wage equations may suffer from measurement error such as mi-reported years of education. IV has also been applied in different contexts. Fish demand function was estimated using IV by Angrist, Graddy and Imbens (2000). Discipline like criminology has also witnessed the application of IV (Angrist, 2006). Econometric issues like errors-in-component have equally been addressed using IV (Arellano & Bover, 1995). Instrumental variable is also extended to a case of binary outcome dependent variable (see for example, Clarke and Windmeijer, 2012). In summary, the method of IV has been used in different contexts with researchers developing and justifying the use of instruments.

The objective of this study is achieved by creating an instrumental variable from the available data and estimating TSLS. The next section reviews the assumptions and the limitations of OLS and presents the strengths of the alternative methods.

Information on the source of data used, diagnostic, hypotheses and estimated models are presented in section three. Section four summarizes the empirical results while section five concludes the study with policy recommendations.

2. Literature Review

OLS has been relied upon for many years as tool for examining the relationship between dependent variable and explanatory variable(s). It however yields an inconsistent estimate when an explanatory variable is endogenous. Some assumptions are central to the Ordinary Least Squares (OLS) and the violation of some of these assumptions may result in a biased and inconsistent estimate. The assumptions underlying OLS are summarized into six as follows.

(i) Linearity assumption: The relationship between the dependent variable and the independent variable are assumed linear.

(ii) The design matrix for the independent variable is of full column rank: Put differently, the independent variables are not linearly related or there is no exact linear relationship among the independent variables in any given linear model.

(iii) The independent variables are assumed to be exogenous: This implies the independent variables should not carry useful information for the prediction of the error term. In other words, there should not be correlation between the independent variable and the disturbance error.

(iv) Homoscedasticity and no autocorrelation: That is, the variance of the error term is assumed constant.

(v) Stochastic or non-stochastic data: The explanatory variable is expected to be random or stochastic.

(vi) Normal distribution of the disturbance error: the error term is assumed to be normally distributed with zero mean and constant variance.

The violation of assumption three, that is when the error term is not equal zero implies that the explanatory variables provide information about the expectations of the disturbances. If this occurs, the OLS estimate will be biased and inconsistent but the IV or TSLS can be used to obtain unbiased and consistent estimate. This is tested in this study. Note that IV and TSLS rely on the above assumptions but includes instrumental variables to correct for endogeneity problem.

2.1 The Models: OLS, IV and TSLS

Ordinary least squares (OLS) forms the basis for econometric analysis while the IV and TSLS extend its specification (Nelson and Startz, 1990). Assuming the data generation process of the equation (1):

$$y_i = \beta x_i + \varepsilon_i \qquad (1)$$

Where i indexes observations, y_i is the observed dependent variable, x_i represents the set of independent variables, ε_i is an unobserved error term representing other factors that may explain the dependent variable other than the included explanatory variables in the model, and β is a scalar parameter to be estimated. If we assume the error terms are serially uncorrelated or normally distributed with a constant variance, given N observation and using the method of moment, the OLS estimator is expressed as equation (2).

$$\hat{\beta}_{OLS} = \beta + \frac{x^N \varepsilon}{x^N x} \tag{2}$$

Where x and εdenote column vectors of the length of the observation. The presence of $\frac{x^N \varepsilon}{x^N x}$ in equation 2 suggests the OLS estimate is biased and inconsistent. In estimating the causal effect of education(X) on wage (Y), an instrument (Z) is a variable which has a direct effect on the explanatory variable but affects the dependent variable indirectly through the endogenous explanatory variable. Assuming X is the $N\ by\ K$ matrix and Z is a$N\ by\ K$ matrix of instruments, the IV estimator of equation (3) provides a consistent estimator.

$$\hat{\beta}_{IV} = (Z^N X)^{-1} Z^N Y \tag{3}$$

Econometric models defining and relating OLS, IV and TSLS are presented next. One important attribute of IV is the identification of the model due to the inclusion of the instrument. A model is said to be exactly identified if the number of instruments equals the number of explanatory variables, over-identified if the number of instruments are more than the number of explanatory variables and under-identified if the number of instruments are less than the number of explanatory variables. Therefore, the order condition for identification is that there must be more instruments than included endogenous variables. Consistent estimate is not possible if the instruments are correlated with the error term and when there is 'weak' instrument.

2.2 The Two-Stage Least Squares (TSLS)

The two-stage least-squares (TSLS) is one of the computational methods used in addressing endogenous problem. Although it has some advantages over OLS, it does not free from limitations. These limitations are well documented in Angrist, Imbens and Krueger (1999). It is worth noting that TSLS gives comparable result to IV especially with one endogenous explanatory variable (Wooldridge, 2002). The TSLS entails estimating an econometric model suspected with an endogenous covariate in two-stages. In the first stage, an endogenous covariate in the equation of interest is regressed on all the exogenous variables in the model, including both exogenous covariates in the equation of interest and the excluded instruments. The predicted value from the first stage regression is obtained and used in the second stage.

Stage 1: Regress each column of Xon Z. That is, estimate equation (4) in the first stage and save the predicted value as in equation (3).

$$X = Z\gamma + \epsilon \tag{4}$$

The first stage estimator is given as: $\hat{\gamma} = (Z^N Z)^{-1} Z^N X$, and save the predicted values of X (equation 5):

$$\hat{X} = Z\hat{\gamma} = Z(Z^N Z)^{-1} Z^N X \tag{5}$$

In the second stage, the regression of interest is estimated as usual, except that in this stage each endogenous covariate is replaced with the predicted values from the first stage.

Stage 2: Regress Y on the predicted values from the first stage (equation 5) as shown in equation 6:

$$Y_{TSLS} = \hat{X}\beta + \epsilon \tag{6}$$

The TSLS estimator,$\hat{\beta}_{TSLS}$(equation 6) is numerically identical to that of IV, $\hat{\beta}_{IV}$(equation 3).

3. Methodology

3.1 Source of Data

This study utilizes the data collected by the National Bureau of Statistics (NBS) in collaboration with the World Bank between August and October 2010 (post-planting) and February and April 2011 (post-harvest). This is the first General Household Survey-Panel (GHS-Panel) to be carried out by the NBS in Nigeria. Five thousands (5000) households were intended to be interviewed from a total number of 500 Enumeration Areas (EAs) in a two-stage stratified sampling. However, the final number of households interviewed was 4,986 out of which 4,946 questionnaires were fully completed. Therefore, data from 4,946 household heads were used in the analysis of this study.

The setting of this study is Nigeria. Nigeria has 774 Local Government Areas, 36 States and the Federal Capital Territory in Abuja with a projected population of over 220 million people by 2050 at a growth rate of 2.5 percent per annum. The country is located in West Africa on the Gulf of Guinea between Benin and Cameroon. She borders Cameroon in the East, Chad in the Northeast, Niger in the North and Republic of Benin in the West. She lies between latitudes $4^{o}1^{1}$and $13^{o}9^{1}$N and longitudes $2^{o}2^{1}$and$14^{o}30^{1}$E. Nigeria covers 923,768 sq km land including water bodies. Nigeria's climatic condition is arid in the North, tropical in the central and equatorial in the South. The temperature ranges between 26 - 36^{o}C in the South and 33 – 40^{o}C in the North. Relative humidity is high during the raining season (between March and November) in the South and (between June and September) in the North while low humidity coincides with the dry season. The annual rainfall decreases Northward averaging 3,550mm in the Niger Delta to 2,200mm in the West and 500-750 mm in the North.

3.2 Specification Tests

There tests are often carried out in the application of IV. This relates to the strength or relevance of the instrument, endogeneity and over-identification restriction or validity of the instrument. The test of instrument relevance involves examining the significant of the Wald statistic or F-statistic. The strength of the instruments can be directly assessed because both the endogenous covariates and the instruments are observable (Stock, Wright and Yogo, 2002). A common rule of thumb for models with one endogenous regressor is that the F-statistic from the model that test the null hypothesis that the excluded instruments are irrelevant in the first-stage regression (equation 4) should be larger than 10.

The Wu-Hausman test, a test of restriction, which is F distributed is used to test for the endogeneity of the explanatory variable. This test is important because TSLS/IV may produce estimates with larger standard errors relative to OLS if an explanatory variable is not endogenous. It is therefore referred to as test of the consistency of OLS. Detailed specifications are presented below.

The third test which tests the validity of instrument, often called Sargan test is Chi-square distributed. This tests over-identification restriction. The most common test of these over-identifying restrictions is based on the observation that the residuals should be uncorrelated with the set of exogenous variables if the instruments are truly exogenous. The Sargan test statistic can be calculated as NR^2 (the number of observations multiplied by the coefficient of determination) from the OLS regression of the residuals onto the set of exogenous variables. This statistic will be asymptotically chi-squared with $m-k$ degrees of freedom (where m equals the number of instruments and k equals the number of exogenous variables) under the null that the error term is uncorrelated with the instruments. Put differently, the

Lagrange multiplier statistic (NR^2) will not exceed the critical point on a χ^2 (r) distribution, where r is the number of over-identifying restrictions ($r = m - k$). Rejection of the null hypothesis casts doubt on the suitability of the instruments.

3.3 Estimated Models for TSLS

The empirical equations estimated in this study are specified from equations (7) to (10) starting with a structural model of equation 7.

$$Y = \beta_0 + \beta_1 X + \beta_2 Z_1 + \beta_3 Z_2 + \mu_i \tag{7}$$

Where:
Y = Wage rate (Naira)
X = Educational level which is assumed endogenous (year)
Wage rate is calculated as amount earned per unit of hours work while education is the number of years of formal schooling. The Instrumental variables used include:

Z_1= Years of experience in non-farm activities

Z_2= Sector (Number of sectors the respondents involved in)

The First Stage Model

The reduced form equation in the first stage is specified as in equation (8):

$$X = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \mu_i \tag{8}$$

This allows the computation of OLS residuals, μ_i.The residuals are not observable but its consistent estimate can be calculated as $\hat{\mu}$. These predicted residuals are included as an additional regressor in the OLS model of equation 9 to test for endogeniety problem That is, when δ is significantly difference from zero. This is called Wu-Hausman test.

$$Y = \beta_0 + \beta_1 X + \delta\hat{\mu} + \varepsilon \tag{9}$$

The Second Stage

Equation (10) is estimated in the second stage model if the null hypothesis of no endogeneity is rejected. An explanatory variable is endogenous if $Cov(\mu_i \mu_j) \neq 0$.

$$Y = \beta_0 + \beta_1 \hat{X} \tag{10}$$

Where: $\hat{X}$= predicted value of the X (education) from the first stage (equation 10)
Y= wage rate as defined in equation 7.

4. Results and Discussion

In this study, different specification tests were carried out as indicated in section three above. First, the endogeniety test is examined followed by the test of instrument strength as well as over-identifying restriction. Finally, TSLS is applied to estimate the return of education on the wage rate among farming households in Nigeria.

4.1 Tests for Instrument Relevance, Endogeneity and Over-identification

The results of the tests for the relevance of the instrument as well as over-identification restriction are presented in Table 1. As noted earlier, the common rule of thumb for model with one endogenous explanatory variable is that F-statistic against the null that the excluded instruments are irrelevant in the first stage regression should be greater than 10.The null hypothesis is that there is no correlation between the explanatory variable and the error term. The R^2 and adjusted R^{-2} are 0.038 and 0.037 respectively while the F-value is 97.22 and is significantly difference from zero at 1%. Following the common rule of thumb, the null hypothesis that the excluded instruments are irrelevant is rejected. Thus, education is an endogenous variable while the instruments (experience and sector) are not only relevance but are important determinants of education. In other words, there is a correlation between year of education and the instrumental variables (experience and sector). Therefore, the instruments are truly exogenous, not correlated with the error term.

Table 1: Test for Instrument Relevance (First Stage OLS Result)

Variable	Coefficient	S.E	t- value	Sig.
Constant	3.734***	0.093	40.15	0.000
Experience	$2.288E^{-5}$	0.000	1.43	0.15
Sector	0.194***	0.014	13.86	0.000
R^2	0.038			
Adj R^{-2}	0.038			
F-value	97.22***			

N=4,946, Dependent variable= education. *** implies significant at 1 percent.
Source: Data Analysis, 2011

The result of the test of endogeniety (equation 9) is presented in Table 2. The OLS estimate gives F-value of 550.433 which is significant at 1%. The residual is significant at 1% level which informs the rejection of the null hypothesis (δ=0) that residuals are not significantly difference from zero. Thus, the instruments are exogenous or the level of education is endogenous suggesting OLS should not be relied on.

One popular test for the validity of the instruments or over-identifying restriction is called Sargan test. This was calculated using the R^2 from the first stage of the regression model (equation 8).The null hypothesis states that the error term is uncorrelated with the instruments$(Cov(\mu, Z) = 0)$.The t-statistic is computed as$NR^2$where N is the number of observation. As earlier noted, this statistic is asymptotically chi-squared with $m - k$ degrees of freedom where m is the number of instruments and k is the number of explanatory variables.

Table 2: Test for Endogeneity (OLS Result)

Variable	Coefficient	SE	t-value	Sig
Constant	2087.49***	368.25	5.67	0.001
Education (X)	-520.03***	96.60	-5.38	0.000
*Residual ($\hat{\mu}$)	499.44***	96.62	5.17	0.000
R^{-2}	0.314			
Adj R^{-2}	0.313			
F-value	550.43***			

N = 4,946, Dependent variable= wage rate. *** implies significant at 1 percent.
Source: Data Analysis, 2011

Comparing the calculated value which is greater than the tabulated value, the null hypothesis of no over-identification is rejected implying there is an over-identification. Indeed, education is the only explanatory variable in the model while there are two instruments. This suggests OLS cannot yield a consistent estimate.

Table 3: The Second Stage Result

Variable	Coefficient	S.E	t-value	Sig
Constant	-1529.406***	47.507	-32.193	0.000
Predicted value of education	439.673***	9.994	43.995	0.000
R^2	0.281			
Adj R^2	0.281			
F-value	1935.57***			

N=4,946, Dependent variable = wage rate. *** implies significant at 1 percent.
Source: Data Analysis, 2011

The result of the second stage (equation 10) is presented in Table 3. The result shows $R^2$and adjusted R^2 values of 0.281 and 0.281 respectively. The F-value is 1935.57 and is significant at 1%. The predicted value of education is positively and significantly (1%) related to wage rate. In addition, the experience and sector of non-farm activities have significant and positive effect on wage rate through direct effect on education.

4.2 Ignoring Endogeneity Problem (Relying on OLS Method)

The result of OLS is presented in Table 4. This test was carried to compare the OLS and TSLS under the assumption that OLS may give a consistent estimator. As shown by the results of test for endogeneity (Tables 1 and 2), the result presented in Table 4 confirms that OLS cannot yield a consistent estimate and should not be used for the estimation, but rather IV(or TSLS). The $R^2$and Adjusted R^2 values are 0.000 while the F-value is 1.854 and not significantly different from zero.

Table 4: OLS Result when Endogeniety Problem is Ignored

Variable	Coefficient	SE	t-value	Sig
Constant	561.32***	12.62	44.46	0.000
Education	-3.122	2.29	-1.36	0.173
R^{-2}	0.000			
Adj R^{-2}	0.000			
F-value	1.85			

N= 4,946, Dependent variable= wage rate. *** implies significant at 1 percent.
Source: Data Analysis, 2011

4.4 Two Stage Least Squares (TSLS) Method: Simultaneous Estimation

The result of the TSLS (running the explanatory variable and the instrumental variables simultaneously) is presented in Table 5. The result shows that R^2 and adjusted R^{-2} values are 0.0032 and 0.032 respectively. The F-value of 162.89 is significantly different from zero at 1%. This result is directly comparable to that obtained in the second stage of the OLS result (presented in Table 2). However, the high F-value of the second stage OLS result as well as higher figures of R^2 and adj R^{-2} attest to the fact that estimate from the two-stages(rather than the simultaneous ones) provides a consistent estimate. Notwithstanding, the estimated coefficients are the same. Indeed, the TSLS result presented in Table 5 further confirms that education is positively and significantly related to wage rate. Thus, the result of the TSLS is unbiased, reliable and consistent. It shows that 1 unit increase in year of education will increase the wage rate by ₦439.67, holding other variables constant. This result agrees with Angrist and Krueger (1991) who found that other variable like quarter of birth has indirect effect on earning in addition to the direct effect of education. It also extend the findings of Angrist and Krueger (2001) used quarter of birth as instrumental variable for ability in wage equation and Angrist and Krueger (1992) who reported parenting as a key omitted variable in a wage equation.

Table 5: TSLS Result (Simultaneous Estimation)

Variable	Coefficient	SE	t-value	Sig
Constant	-1529.67***	163.758	-9.339	0.000
Education	439.67***	34.449	12.76	0.000
R^{-2}	0.032			
Adj R^{-2}	0.032			
F-value	162.89***			

N= 4,946, Dependent variable = wage rate. *** implies significant at 1 percent.
Source: Data Analysis, 2011

5. Conclusion and Policy Implications

Testing the potential endogenous problem of an independent variable is important in econometrics analysis and economic application because it gives information on the appropriate estimation method to be used as well as guiding policy decisions. For instance, relying on the classical least squares (OLS) when regressor is endogenous yields a biased and inconsistent estimator as shown in this study. However, IV or TSLS gives an unbiased and consistent estimate when the independent variable is endogenous. This suggests the use of instruments or instrumental variables. Three tests are important when instruments are employed in TSLS or IV estimation. These include test of the validity or strength of instruments, test of endogeneity as well as test of over-identifying restrictions. The null hypotheses are rejected for all the three tests in this study suggesting TSLS or IV is desirable. In other words, it was demonstrated in this study that OLS provides an inconsistent estimate. More importantly, the empirical results show that education is an important determinant of the wage received by farm households in Nigeria. In conclusion, economic researchers, social scientists and development experts are advised to ascertain the potential endogenous problem associated with the independent variables as a guide to identifying the appropriate econometric estimation method to apply. This has significant effects on the estimates of econometric models and subsequently policy suggestions (inference) emanating from such estimates. For example, the return to education is not zero as revealed by the OLS result. Finally, investment in education as a veritable tool for wage increase would have multiplier effects on economic development in Nigeria.

References

Angrist, J. D. (2006). Instrumental Variables Methods in Experimental Criminological Research: What, Why and How. Journal of Experimental Criminology, 2(1): 23-44.

Angrist, J. D. & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4): 69–85.

Angrist, J. D., Graddy, K. & Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies.* 67 (3): 499-527

Angrist, J. D. Imbens, G. W. & Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14 (1): 57-67.

Angrist, Joshua D., Imbens, Guido W. & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91 (434): 444-455.

Angrist, J. D. & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90 (430): 431-442.

Angrist, J. D. & Krueger, A. B. (1992). The effect of age of school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87 (418): 328-336.

Angrist, J. D. & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106 (4): 979-1014.

Arellano, M. & Bover O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68 (1): 29-51.

Balke, A. & Pearl, J. (1997).Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1172–1176.

Bound, J. Jaeger, D. A. and Baker, R. M. (1995).Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, 90 (430): 443-450.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, But What's the Mechanism? (Don't Expect an Easy Answer). *Journal of Personality and Social Psychology*. 98: 550-58.

Clarke, P. S. & Windmeijer, F. (2012). Instrumental variable estimators for binary outcomes. *Journal of the American Statistical Association*, 107 (500): 1638-1652.

Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources*, 32(3): 441–462.

Imbens, G. & Angrist, J. (1994).Identification and estimation of local average treatment effects, *Econometrica*. 62: 467–476.

Leigh, J. P. & Schembri, M. (2004). Instrumental variables technique: cigarette price provided better estimate of effects of smoking on SF-12.*Journal of Clinical Epidemiology*. 57(3): 284–293.

Miguel, E., Satyanath, S. & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variable approach. *Journal of Political Economy*. 112: 725–753.

Nelson, C. R., & Startz, R. (1990). Some further results on the small sample properties of the instrumental variable estimator. *Econometrica*. 58(4): 967–976.

Stock, J.H., Wright, J. & Yogo, M. (2002). A Survey of weak instruments and weak identification in Generalized Method of Moments, *Journal of the American Statistical Association*, 20(4): 518–29.

Stock, J. H. & Trebbi, F. (2003). Retrospectives: Who Invented Instrumental Variable Regression? *Journal of Economic Perspectives.* 17 (3): 177–194.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*: MIT press, Cambridge, Massachusetts, London, England, 881 pp.