Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Addressing Endogeneity in Empirical Research: A Review of Instrumental Variable Techniques

[1]Mrs. Nancy Purohit, [2]Dr. Tushar Kanti Das
[1]Research Scholar, Department of Business Administration, Sambalpur University, Jyoti Vihar,Burla,Odisha
[2]Professor, Department of Business Administration, Samblapur University, Jyoti Vihar, Burla, Odisha

ARTICLE INFO

Received: 18 Oct 2024

Revised: 10 Nov 2024

Accepted: 28 Dec 2024

ABSTRACT

In order to determine the efficacy of public policy interventions meant to improve socioeconomic outcomes, impact evaluation of welfare programs is essential. However, the existence of endogeneity resulting from non-random selection, omitted variable bias, reverse causality, and measurement errors frequently complicates empirical estimation of such impacts. To overcome these difficulties and find causal relationships in observational data, the Instrumental Variable (IV) approach has become a potent econometric method. The theoretical underpinnings, statistical framework, presumptions, and applications of IV methods in impact evaluation are all thoroughly reviewed in this paper. Additionally, it provides a clear example of IV estimation, critically assesses its advantages and disadvantages, and goes into great detail about the problem of endogeneity. The study emphasizes how crucial it is to choose reliable instruments and carry out robustness tests in order to guarantee reliable inference. The review concludes that although IV methods have significant benefits in causal analysis, their efficacy is largely dependent on the veracity of underlying assumptions and empirical application.

Keywords: Instrumental Variable (IV), Endogeneity, Impact Evaluation, Omitted variable bias, socio economic outcomes.

INTRODUCTION:

In developing economies, where welfare programs are used to combat poverty, unemployment, and inequality, impact evaluation is essential to policy analysis. The goal of programs like income support policies, social protection programs, and employment guarantee schemes is to raise household welfare metrics like savings, consumption, and income.

However, the lack of randomised assignment makes estimating the causal impact of such programs intrinsically difficult. In the majority of real-world situations, eligibility requirements or voluntary participation in welfare programs result in selection bias. Participating households may exhibit systematic differences from non-participating households in both observable and unobservable aspects, including financial behaviour, risk preferences, and motivation.
When the explanatory variables are endogenous, conventional econometric methods such as Ordinary Least Squares (OLS) are unable to take these problems into consideration. Estimated coefficients may therefore be erratic and biased. The Instrumental Variable (IV) approach, which isolates exogenous variation in the endogenous regressor to enable consistent estimation, is being used more and more by researchers to get around this restriction.

In addition to discussing their use in welfare economics research, this paper examines the conceptual and empirical significance of IV methods in impact evaluation.

Endogeneity in Impact evaluation:

Normal regression is where we estimate impact is :

$$y_i = \beta_o + \beta x_{1i} + u_i$$

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Where ,

y_i = Outcome/Dependent variable for Individual "i"

β_0 = Intercept

β_1 = Coefficient of Independent Variable

x_{1i} = Independent variable /Explanatory (Endogenous)

u_i = Error term

When an explanatory variable and the error term are correlated, endogeneity—a major issue in econometric analysis occurs where,

$$\text{Cov}(x1i, ui) \neq 0 : x_{1i} \text{ is endogenous.}$$

This results in skewed estimates and goes against the traditional exogeneity assumption where normal OLS can't be applied to estimate the impact.

Endogeneity Sources:
1. Ignored Variable Bias
The model excludes unobserved variables that have an impact on both the independent and dependent variables.
2. Causality in reverse
The dependent variable may be causally related to the independent variable.
3. Inaccurate Measurement
Correlation with the error term may be introduced by inaccurate variable measurement.

Problem of endogeneity: An illustration:

Endogeneity issue is prevalent in many impact evaluation study, for example a well known social welfare scheme in India is MGNREGA (Mahatma Gandhi National Rural Employment Guarantee Act),

MGNREGA participation is more common among low-income households. Even in the absence of the program, low-income households typically have different savings habits.Unobserved or pre-existing factors influencing household savings are correlated with the variable MGNREGA participation. Participation is therefore not exogenous. Why this is an issue ,if we perform a straightforward comparison like this: participants as opposed to non-participants, or OLS regression of participation-related savings. MGNREGA's estimated impact could be skewed. It's possible that poverty status, rather than the program itself, accounts for some of the observed savings disparity.
In terms of econometrics

This leads to endogenous participation and selection/self-selection bias because:

Cov(MGNREGA participation, error term) ≠ 0

The error term could include elements linked to poverty such as:

1)financial susceptibility

2)low base of assets

3)Erratic income

4)Absence of alternative work

5)Risk-coping strategies

They affect savings as well as participation. A method that takes this endogeneity into account is required. In that case, 2SLS/Instrumental Variable (IV) useful if a legitimate instrument that influences participation but not savings directly, and there is also unobserved selection.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Instrumental Variable

When explanatory variables are endogenous, an instrumental variable is a tool for obtaining reliable estimates. A legitimate instrument needs to meet:

A valid instrument must satisfy following 2 conditions:

1) Relevance Condition:

$$Cov(Z_i, X_i) \neq 0 \text{ (Where } Z_i = \text{Instrument}, X_i = \text{Independent Variable)}$$

2) Exogeneity Condition:

$$Cov(Z_i, u_i) = 0 \text{ (Where } u_i = \text{Error term)}$$

The IV method eliminates correlation with the error term by substituting the endogenous variable with its predicted value obtained from the instrument.

IV Regression: Statistical Equation:

Structural Equation:

$$Y_i = \beta_0 + \beta_1 x_1 + u_i, \text{ (} x_1 \text{ is endogenous)}$$

First stage Equation :

$$X_{1i} = \pi_0 + \pi_1 + V_i$$

Let, $\pi_0 + \pi_1 = y_2$,

So, $X_{1i} = y_2 + v$,where, y_2 is called the systematic component and V is non systematic.

Second stage Equation:

y_2 will be estimated as $\hat{y}_2$

$$\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1$$

Then putting the value of $\hat{y}_2$ in the original structural form equation,

$$Y_i = \beta_0 + \beta_1 \hat{y}_2 + \beta_3 V + u_i \text{ , where } \beta_3 V + u_i \text{ is the composite error term.}$$

This method is also termed as 2stage least square method. (2SLS)

Assumption of Instrumental Variable Method:

1. Significance
The endogenous variable must be strongly predicted by the instrument.
2. Exogeneity
The error term cannot be correlated with the instrument.
3. Limitations on Exclusion
The endogenous variable is the only way the instrument influences the dependent variable.
4. Monotonicity
Every person is impacted by the instrument in the same way.

Instrumental Variable (IV) Method Benefits:

1. Offers reliable estimates in the event of endogeneity:

When an explanatory variable and the error term are correlated, IV corrects bias. It provides accurate estimates by isolating only the exogenous variation.

2. Makes causal interpretation possible:

Instead of just finding correlations, IV assists in identifying cause-and-effect relationships. When randomized experiments are not feasible, this is particularly helpful.

3. Unobserved heterogeneity controls

It takes into consideration hidden variables that influence both variables, such as ability and poverty level.

As a result, bias from missing variables is lessened.

4. Useful in a variety of fields

Research in economics, health, education, and public policy all make extensive use of IV. It is adaptable and helpful in numerous empirical studies conducted in the real world.

Instrumental Variable (IV) Method Drawbacks:

1. Having trouble locating reliable instruments

A good tool should influence participation without having a direct impact on the result. These variables are uncommon and frequently difficult to theoretically support.

2. A weak instrument issue

Results become untrustworthy if the instrument and the endogenous variable have a weak correlation. Large standard errors and skewed estimates may result from this.

3. Calculates the Local Average Treatment Effect (LATE)

IV only records the effect for a particular subgroup that the instrument has an impact on. It might not accurately reflect the effects on the whole population.

4. Receptive to specifications

The model setup and instrument selection have a significant impact on the results. Different conclusions can result from minor modifications to the specification.

CONCLUSION:

An essential econometric method for handling endogeneity problems in impact evaluation studies is the Instrumental Variable (IV) approach. It allows researchers to go beyond simple correlations and derive significant causal relationships by taking advantage of exogenous variation, especially in non-experimental settings like welfare program analysis. Because of this, IV is particularly useful for assessing programs like MGNREGA, where participation is determined by underlying socioeconomic factors rather than being assigned at random.

Despite its advantages, the validity and strength of the selected instrument have a significant impact on the IV method's efficacy. The reliability of the findings can be compromised by biased and inconsistent estimates produced by a weak or invalid instrument. Thus, a good IV strategy must include rigorous empirical testing, such as relevance and exogeneity checks, careful instrument selection, and solid theoretical justification.

It may be concluded that, despite the complexity involved in the application of the IV approach, it remains an indispensable tool in the field of policy evaluation research. As long as the IV approach is applied correctly, it promises to deliver reliable and valuable insights that may be useful in the formulation of policies. As far as the application of the IV approach in the field of welfare policies, including the MGNREGA, is concerned, it may be said that the application of the IV approach helps in the determination of the reliability of the impact estimates, thus promoting the effectiveness of the policies formulated in the field

REFERENCES:

[1] Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.
[2] Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.
[3] Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
[4] Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.
[5] Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

[6] Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification. *Journal of Economic Perspectives*, 15(4), 69–85.
[7] Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
[8] Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444–455.
[9] Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467–475.
[10] Angrist, J. D., & Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of Labor Economics* (Vol. 3A).
[11] Wooldridge, J. M. (2008). Instrumental variables estimation of the average treatment effect. *Advances in Econometrics*, 21, 93–116.
[12] Wooldridge, J. M. (2016). Should instrumental variables be used as matching variables? *Research in Economics*, 70(2), 232–237.
[13] Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3), 669–738.
[14] Card, D. (2001). Estimating the return to schooling. *Handbook of Labor Economics*, 3, 1801–1863.
[15] Bascle, G. (2008). Controlling for endogeneity with instrumental variables. *Strategic Organization*, 6(3), 285–327.
[16] Semadeni, M., Withers, M. C., & Certo, S. T. (2014). The perils of endogeneity. *Strategic Management Journal*, 35(7), 1070–1079.
[17] Wilms, R., et al. (2021). Omitted variable bias and causal estimation. *Econometrics Journal*, 24(1), 1–20.
[18] Pearl, J. (2012). On a class of bias-amplifying variables. *arXiv preprint arXiv:1203.3503*.
[19] Swanson, S. A., & Hernán, M. A. (2014). Instrumental variables perspective. *arXiv preprint arXiv:1410.0477*.
[20] Levis, A. W., Kennedy, E. H., & Keele, L. (2024). Nonparametric IV estimation. *arXiv preprint arXiv:2402.09332*.
[21] Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM. *Stata Journal*, 3(1), 1–31.
[22] Stock, J. H., & Yogo, M. (2005). Testing for weak instruments. In *Identification and Inference*.
[23] Sargan, J. D. (1958). Estimation using instrumental variables. *Econometrica*, 26(3), 393–415.
[24] Hansen, L. P. (1982). Large sample properties of GMM estimators. *Econometrica*, 50(4), 1029–1054.
[25] Keane, M. P., & Neal, T. (2024). A practical guide to weak instruments. *Annual Review of Economics*, 16, 185–212.
[26] Imbert, C., & Papp, J. (2015). Labor market effects of MGNREGA. *American Economic Journal: Applied Economics*, 7(2), 233–263.
[27] Zimmermann, L. (2020). Why guarantee employment? Evidence from MGNREGA. *Journal of Development Economics*, 145, 102428.
[28] Afridi, F., Mukhopadhyay, A., & Sahoo, S. (2016). Female labor force participation and MGNREGA. *World Development*, 83, 123–138.
[29] Berg, E., Bhattacharyya, S., Durgam, R., & Ramachandra, M. (2012). Can rural public works affect agricultural wages? *World Bank Economic Review*, 26(2), 192–213.
[30] Narayanan, S., & Das, U. (2014). Women participation in MGNREGA. *World Development*, 61, 1–17.
[31] Ravallion, M. (2007). Evaluating anti-poverty programs. *Handbook of Development Economics*, 4, 3787–3846.
[32] Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in econometrics. *Journal of Economic Literature*, 47(1), 5–86.