Instrumental variables in psychology

Hugo Hanema
Utrecht University
h.hanema@students.uu.nl

ABSTRACT
In the social sciences, it is often of great importance to clearly uncover causal relationships. However, many researchers are unaware of the issue of endogeneity which biases estimates of causal effects. The instrumental variable (IV)-approach solves endogeneity and can make a convincing argument for causality even with cross-sectional data, but has been under-utilized in the social sciences. This paper explains the IV-approach and provides an example of application of this method to the psychological research question of causality between self-esteem and depression. The main argument is that the IV-approach is applicable to and deserves to see more use in psychology.

Keywords
Causality, endogeneity, instrumental variables.

INTRODUCTION

The term 'causality' indicates that one event is the result of the occurrence of another event. Statistically, we speak of causality when the value of a dependent variable *y* changes as a result of a change in the value of an independent variable *x*. When attempting to design an intervention to produce a desired effect or prevent an undesired effect, it is indispensable that a causal relationship is clearly uncovered first. If not, the risk exists that much time and effort are spent unfruitfully on designing unsuccessful interventions.

Ideally, questions of causality are studied in an experimental setting. However, it requires little explanation as to why, from ethical and practical standpoints, this is often not an option in the social sciences. So researchers have resorted to different methods of empirically assessing causality. One commonly employed method is longitudinal regression. This method is based on the logic that causes must precede consequences temporally, so when there are serval moments of measurement we can identify the cause and the consequence. Longitudinal regression may, under certain circumstances, be a suitable method to infer causality, but from a practical standpoint longitudinal studies are not preferred. They are more complicated to conduct, more financially demanding and more labor-intensive. In many instances, longitudinal data is simply not available. Regular, cross-sectional regression analysis theoretically is able to infer causality, but only under certain circumstances that in practice are almost never met. So a method which could study questions of causality using cross-sectional data and which is unaffected by the problems of regular, cross-sectional regression analysis would provide much benefit to researchers in the social sciences.

Fortunately, a method that fits these criteria exists, namely the instrumental variable (IV)-approach. The IV-approach has been proposed as a solution to a number of diverse problems which plague researchers who work with non-experimental, observational data (Bollen, 2012). Though diverse, all of these problems have in common that they involve endogeneity, meaning that a correlation exists between one or more explanatory variables and the error term of *y*. This biases and possibly even invalidates estimates from regression analysis. The IV-approach can solve the issue of endogeneity. In addition, it provides a more convincing argument of causality than regular regression analysis when working with cross-sectional, observational data since it relies on exogenous variance in the independent variable. This means that a valid instrument induces change in an independent variable, but does not affect the dependent variable in any way other than through this effect on the independent variable. As such, a causal effect of the independent variable on the dependent variable can be made clear.

Despite its obvious advantages, the IV-approach may still be described as "under-utilized" (Bollen, 2012), at least within the social sciences. IV's have seen regular use in the fields of economics, epidemiology and political science. However, knowledge and use of IV's is much more limited in sociology and certainly in psychology.

The contribution of this paper is that it will provide both an explanation of the IV-approach and an example of application of this technique to the psychological research question of the causal relationship between self-esteem and depression. My main purpose will not be to answer this question of causality since strong evidence has already been provided by Sowislo & Orth's (2013) meta-analysis, but rather to test whether it would be possible to reach similar conclusions as Sowislo & Orth while using cross-sectional data.

In the following section, I will start by describing the most commonly used method to infer causality, regular regression analysis, and make clear the flaws of this method. After that I will explain what the IV-approach is and how it can solve the endogeneity issue and make a convincing argument for causality. In the third section, I will provide an example of application of the IV-approach to the psychological research question. I will highlight the difference in results between regular regression analysis and the IV-regressions when applied to the same dataset. In the fourth section I will describe the limitations of the IV-approach and with these limitations in mind I will reflect upon the results that I obtained. Concluding, I will make a suggestion for further use of the IV-approach in psychology.

CAUSALITY AND THE IV-APPROACH

Regression analysis

In the social sciences, regression analysis is the most commonly used way to assess causality. It is an application of the general linear model (GLM; Tabachnick & Fidell, 2001). A general regression equation with multiple predictors may be written as:

$$y = \alpha + \beta x_1 + \beta x_2 + ... + \varepsilon \qquad (1)$$

where *y* represents a continuous dependent variable that is predicted by the explanatory variables and/or covariates x_1, x_2, ... The constant α represents the value of *y* when all predictors equal zero and the regression coefficient β represents the change in y for a change of 1 in x. The error term ε represents the residual variance: the discrepancy between the observed values of y and the values that are predicted by the equation. The error term includes random measurement error, plus all other influences on y that are not explicitly included in the equation.

When assessing causality using regression analysis, two conditions must be met. Firstly, the direction of the effect must be clear: in other words, there must be some way to determine that x causes y and it is not actually y that causes x. One way to determine this is through longitudinal regression. Another way is to theoretically argue what the direction of the effect must be. However, theoretical arguments often fail to give absolute certainty regarding the direction of the effect and longitudinal studies are not always feasible. Secondly, regression analysis assumes that none of the variables in the model are endogenous – an assumption that will be broken in a variety of situations. One of these situations is the omission of a confounding variable, so another condition to infer causality, that there must be no confounding variables, falls under the condition that there must be no endogeneity.

The problem: endogeneity

What is endogeneity?

An endogenous variable is one that is affected by other variables in the equation via a path that is not accounted for in the model. In contrast, this is not the case for an exogenous variable. Only exogenous variables can provide us with valid estimates. To better understand the problem of endogeneity, let us start by looking at two equations that express the assumption of exogeneity:

$$E(\varepsilon) = 0 \tag{1}$$

and

$$COV(x, \varepsilon) = 0 \tag{2}$$

Equation 1 shows that we expect the error term to vary randomly and therefore have a mean of zero taken across all cases. There are multiple situation in which this assumption will be broken, but for our purposes we will focus on the reason which relates to Equation 2. This equation is the formal definition of exogeneity. It shows that no correlation is assumed to exists between any of the predictors and the variables that are included in the error term. When such a correlation exists, the error term will not have a mean of zero because different cases will have different values for this predictor which correlates with the error term so the error term will vary in a nonrandom manner. The regression equation is based on the assumption that the error term has a mean of zero, so when in actuality it has values that it is assumed not to have, all other unknown parameters in the equation will be estimated with bias.

There are several situations that give rise to a correlation between a predictor and the error term. As mentioned, one such situation is when a confounding variable is omitted. Another situation is that of a feedback relation. We speak of a feedback relation when x causes y, but the reverse is also true. Consider for example that people's smoking habits are party determined by social influence (i.e., the presence of other smokers). This means that my smoking behavior (y) may be caused by my friend's smoking behavior (x), but the reverse is equally true. Figure 1 illustrates our model when we attempt to estimate y from x while a feedback relation is present. Dotted lines represent relations that exist in actuality, but aren't accounted for in the model.

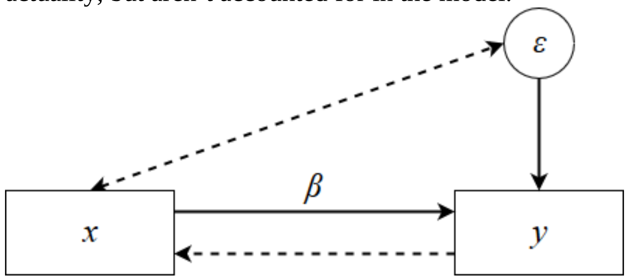

Figure 1. Path diagram of predicting y from x when a feedback relation exists.

When a feedback relation exists between x and y, this means that there is a path from y to x that is not accounted for in the model. Consider that the error term of y includes all variables that affect y but are not explicitly included in the model. Since y has a causal effect on x, it also has an effect on itself through its effect on x. This effect of y on itself is not accounted for in the model, so y becomes a part of its own error term! Figure 1 illustrates the model where y is the dependent variable, but the same logic applies when we estimate x from y (which we should do when we suspect that a feedback relation exists because we must model both the path from x to y and from y to x to prevent biased estimates). Following the same line of reasoning, x is part of its own error term, too. So in summary: x and y are related and y is part of its own error term, so x and ε_y are related as well. Since the variables have a reciprocal effect, y and ε_x are also related.

The solution: instrumental variables

What are instrumental variables?

The core idea of the IV-approach is that the problem of $COV(x, \varepsilon) \neq 0$ can be circumvented by substituting the observed values of x in the regression equation with estimated values of x, given that these values are estimated by one or more instrumental variables which are completely unrelated to y. An instrumental variable – suppose we call it z – must satisfy two main conditions (Woolridge, 2015):

$$COV(z, \varepsilon) = 0 \tag{3}$$

and

$$COV(z, x) \neq 0 \tag{4}$$

As expressed in these equations, the first condition is that z is uncorrelated to the error term. The second condition is that z must be correlated to x. If this were not the case, it would be impossible to estimate x from z.

Application of IV's follows the two steps of the Two Stage Least Squares (2SLS) procedure. The first step is the first-stage regression: the regression to estimate x from our instruments. Then comes the second stage of the 2SLS-procedure, which is to estimate our original regression model except now using the previously estimated values of x. By substituting x as it is observed for an estimate of x based on our instruments, we are left only with the exogenous part of variance in x. From this, we can estimate y without bias. In the case of a feedback relation, the reciprocal causal effect still exists, but it is no longer problematic because now y is predicted from an estimate of

x which is unrelated to the error term of y. The same but vice versa is true when predicting x from y.

The IV-approach's ability to infer causality lies in the fact that IV's create exogenous variance in an explanatory variable – in other words, the fact that IV-estimation produces values of x that are completely unaffected by y. Because of this, firstly we can be sure of the direction of the effect: since z is exogenous, it follows that x as it is estimated by z is also exogenous and so is per definition unaffected by y. This leaves us with only one possible direction of the effect: from x to y. Secondly, as discussed, when using IV's we can be sure that there is no endogeneity. This both guarantees that there are no omitted confounding variables and that the causal effect is estimated without bias.

APPLICATION OF THE IV-APPROACH

Causality between self-esteem and depression

Self-esteem and depression are two constructs that are known to be strongly related, though until relatively recently little was known about their prospective effects on each other (Orth, Robins & Roberts, 2008). Concerning this matter, there are four possibilities. The first possibility is that having low self-esteem causes a person to become depressed, a hypothesis that is crystallized in the *vulnerability model*. The second possibility is that suffering from depression causes a person to have decreased self-esteem. The model that corresponds to this hypothesis is the *scar model*. The vulnerability model and scar model are not mutually exclusive: a third possibility is that they both are true and self-esteem and depression have a causal effect on each other and therefore would be said to have a feedback relation. The fourth possibility is that none of the above possibilities are true and self-esteem nor depression have a causal effect on the other. Sowislo & Orth (2013) have conducted a meta-analysis of longitudinal studies and concluded that self-esteem and depression cause each other, however the causal effect of self-esteem on depression is much stronger so their results may be interpreted as providing strong evidence for the vulnerability model and only weak evidence for the scar model. Based on this, I hypothesize that self-esteem and depression form a feedback relation and from this endogeneity arises, so regular regression will provide biased estimates. In addition, regular regression analysis cannot guarantee that there are no confounding variables in the relationship between self-esteem and depression. As an illustration of the IV-approach applied to a psychological research question, I will now estimate both the causal effect of self-esteem on depression and vice versa using instrumental variables.

My study

The data

I used publicly available data from the National Longitudinal Survey of Youth (NLSY79), a US longitudinal panel-survey. I only used data from one specific moment of measurement so my dataset is cross-sectional. The sample of young adults I used is estimated to represent 90% or their ago cohort (Center for Human Resource Research, 2006). Due to budgetary reasons, the NLSY79 contains a complex pattern of missing data: by far not all of the total included variables were measured for every respondent in every assessment. Because of this, I did not use the total dataset of 11,521 respondents but instead a subsample of 646 respondents. For the subsample, respondents' ages range from 15 to 38, $M = 20$, $SD = 5.52$. The subsample includes 355 males and 291 females. Self-esteem was measured by the Rosenberg Self-Esteem scale (RSE) and depression by the Center for Epidemiological Studies Depression Scale (CES-D), which are both commonly used, well-validated and reliable measures. In order to avoid omitting any confounding variables, I added a number of covariates to the model which affect both self-esteem and depression.

Regression analysis

First, I estimated both the causal effects of self-esteem on depression and vice versa using regular regression analysis. When self-esteem was entered into the model as dependent variable, depression significantly predicted self-esteem, $\beta = -.11$, $t(625) = -3.81$, $p < 0.001$. R^2 of the model was .515. This result would indicate that depression causes self-esteem, thus confirming the scar hypothesis. When depression was entered into the model as dependent variable, self-esteem significantly predicted depression, $\beta = -.20$, $t(625) = -3.81$, $p < 0.001$. R^2 of the model was .356. This result would also confirm the vulnerability hypothesis. If it were not for the considerations that have been described in this paper, we would conclude that self-esteem and depression have reciprocal causal effects.

The IV-approach

Next, I conducted IV-regressions to see how they would compare to the regular regression analysis. I started with assessing the causal effect of depression on self-esteem. The instruments I used for depression are the average number of hours of sleep that the respondents report to get on a typical weeknight, and two particular items of the CES-D, the first and the fifth, which explain variance in depression but not in self-esteem. Recall from the previous section that the IV-approach follows the Two Stage Least Squares procedure, the first stage being the regression to predict the endogenous regressor – in this case depression – from the exogenous instruments. The first-stage regression explained a significant proportion of variance in depression, $F(22, 623) = 36.79$, $p <.001$, $R^2 = .565$. For the second stage of the Two Stage Least Squares procedure, we run a regression as usual except now using the previously estimated, exogenous values of depression. Depression did not significantly predict self-esteem, $\beta = .000$, $z = .04$, $p = .971$. In contrast to the estimate from the regular regression, from this estimate we would conclude that depression does not have a causal effect on self-esteem and we would reject the scar model.

I proceeded by assessing the causal effect of self-esteem on depression. The instruments I used for self-esteem are one item of the Rosenberg Self-Esteem Scale which explains variance in self-esteem but not in depression, two items of the Pearlin Mastery Scale and one item of a scale which measures the respondent's tendency to engage is risky behaviors. The first-stage regression explained a significant proportion of variance in self-esteem, $F(22, 623) = 50.73$, $p < .001$, $R^2 = .642$. Self-esteem did not significantly predict depression, $\beta = .179$, $z = 1.76$, $p = .078$. From this result we might conclude that, in contrast to regular regression, the IV-regression indicates that self-esteem does not have a causal effect on depression so we would also reject the vulnerability model. Alternatively, we might conclude that there exists a causal effect of self-esteem on depression but it failed to reach statistical significance because IV-regressions inevitably sacrifice some accuracy in its estimates which manifests in larger standard errors. In the regular regression analysis, self-esteem had a standard error of .053 but in the IV-regression this standard error nearly doubled to .101. So it could be argued that the result does provide evidence for a causal effect, however then the positive coefficient would lead us to the counterintuitive conclusion that having higher self-esteem causes a person to become more depressed, so I hypothesize that a sign change has occurred. Whichever of these two interpretations is true, it is certain that the IV-regression points to a different conclusion than the regular regression analysis: either there is no causal effect of self-esteem or there is a causal effect but it is overestimated in regular regression.

DISCUSSION

General limitations of the IV-approach

The quality of an IV-regression is only as high as that of the instruments. An instrument that is only weakly related to x is referred to as a ‘weak’ instrument. Weakness of instruments leads to inconsistency in the IV-estimates. A ‘bad’ instrument is one that is not truly uncorrelated to the error term. When using bad instruments, endogeneity still exists. Since variables included in the error term are unobserved, we cannot simply check if a correlation exists. The diagnostics to test whether $COV(z, \varepsilon) = 0$ that do exist are well-known to be inconsistent. In short, the main pitfall of the IV-approach is that it can be very difficult to find valid instruments.

Results and limitations of my study

I argue that my result regarding the lack of a causal effect of depression on self-esteem is valid. According to the literature, some causal effect of depression on self-esteem likely exists but it is so weak that we would only expect to find a significant effect in studies with major statistical power such as meta-analyses. Judging by findings from previous research, the IV-regression I conducted reflects the truth more so than regular regression using the same data and covariates. Therefore I believe that I have successfully demonstrated that the IV-approach can and should be applied to psychological research questions which involve an endogeneity issue. The result I found regarding the causal effect of self-esteem on depression is more difficult to interpret. The interpretation that self-esteem does not have a causal effect on depression is incongruent with previous research. If we concluded that a causal effect does exist though it did not reach statistical significance due to an enlarged standard error, we still have to explain the positive coefficient. In regression, there are many reasons why a sign change might occur: Kennedy (2005) lists 19. It is intriguing that regular regression did produce a negative coefficient, so I speculate that the issue must lie in the estimate of self-esteem from the first-stage regression. However, since I lacked the necessary time and expertise to check which of Kennedy’s reasons might be applicable to my analysis I am not in a position to draw a definitive conclusion on how the positive coefficient came to be.

I believe that there are two main limitations to my study. Firstly, I cannot be completely certain that the instruments I used were truly uncorrelated to the error term. I did not formulate theoretical arguments as to why this should be the case and the relevant diagnostics cannot provide definite evidence. The second limitation relates to generalizability. Though the total sample is estimated to represent 90% of the population, due to missing values I only used a relatively small subset of this sample. This is a threat to generalizability because of the possibility that the data is missing according to some nonrandom pattern.

Conclusion

In contrast to the regular regression, the IV-regression I conducted showed no causal effect of depression on self-esteem. I argue that this finding is concurrent with previous research and demonstrates that researchers in psychology can and should use the IV-approach to deal with endogeneity issues. Simultaneously, the result I obtained regarding the causal effect of self-esteem on depression shows that the IV-approach does have its difficulties: enlarged standard errors make it harder to find statistically significant results, and β-coefficients may unexpectedly switch between being positive or negative. Generally, it is a challenge to find valid instruments. Still, all taken together I conclude that the IV-approach deserves to see more use within the field of psychology.

ROLE OF THE STUDENT

Hugo Hanema was an undergraduate psychology major writing his bachelor thesis at the Utrecht University department of Methodology & Statistics under the supervision of Lion Behrens and Irene Klugkist. The topic of instrumental variables was proposed by Lion Behrens; the student thought to apply this method to the psychological research question and independently gathered all the necessary data and conducted the analyses. The supervisors assisted by providing advices regarding the more technical methodological and statistical details.

REFERENCES

1. Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. Annual Review of Sociology, 38, 37-72.
2. Center for Human Resource Research. (2006). NLSY79 child and young adult data users guide.
3. Kennedy, P. E. (2005). Oh no! I got the wrong sign! What should I do?. The Journal of Economic Education, 36(1), 77-92.
4. Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. Journal of personality and social psychology, 95(3), 695.
5. Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta - analysis of longitudinal studies.
6. Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). Using multivariate statistics.
7. Wooldridge, J. M. (2015). Introductory econometrics: A modern approach. Nelson Education.