Proving causal relationships using observational data

Henry L. Bryant
Texas A&M University
h-bryant@tamu.edu

David A. Bessler
Texas A&M University
d-bessler@tamu.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 20111

Proving causal relationships using observational data

Henry L. Bryant
Texas A&M University

David A. Bessler
Texas A&M University

April 27, 2011

Abstract

We describe a means of rejecting a null hypothesis concerning observed, but not deliberately manipulated, variables of the form H_0: $A \not\rightarrow B$ in favor of an alternative hypothesis H_A: $A \rightarrow B$, even given the possibility of causally related unobserved variables. Rejection of such an H_0 relies on the availability of two observed and appropriately related instrumental variables. While the researcher will have limited control over the confidence level in this test, simulation results suggest that type I errors occur with a probability of less than 0.15 (often substantially less) across a wide range of circumstances. The power of the test is limited if there are but few observations available and the strength of correspondence among the variables is weak. We demonstrate the method by testing a hypothesis with critically important policy implications relating to a possible cause of childhood malnourishment.

1 Introduction

A burgeoning literature describes the inference of causal relationships among observed random variables when controlled experiments are not conducted (Glymour and Cooper, 1999; Pearl, 2000; Spirtes et al., 2000). This literature generally documents the development of algorithmic approaches to inferring causality among a possibly large number of variables. Such algorithms facilitate an exploratory approach to causal inference, wherein the researcher does not explicitly form or test specific causal hypotheses.

A nascent offshoot from this literature, however, applies the logic that underlies such algorithms to the problem of investigating specific causal hypotheses. Bryant et al. (BBH, 2009) demonstrate how, under certain conditions, one can use observational data on three variables, say A, B and C, to conclude, in the presence of latent variables or not, that A does not cause B. They use Monte Carlo methods to demonstrate reliable small sample properties of such inference. The key idea is that a non-zero correlation between A and B can be interpreted as causal and not just associational under certain reasonable assumptions, and that causality cannot run from A to B if there exists a third variable C (call this an instrument) such that A and B are correlated, A and C are correlated, and C and B are not correlated. Abstracting from the possibility of causally related latent variables, the associated causal structure is $C \rightarrow A \leftarrow B$. Thus, an "inverted fork" in the causal structure is the revealing feature that facilitates inference of the direction of causal flow.

BBH describe two avenues by which a hypothesis that A causes B might be rejected, which they label weak-basis rejections and strong-basis rejections. The former rejections arise when the correlation between A and B is judged to be not significantly different from zero (on the basis of a statistical test). Strong-basis rejections arise from rejection of the hypothesis that the correlation between A and B is zero, rejection of the hypothesis that the correlation between A and C is zero, and failure to reject the hypothesis that the correlation between C and B is zero. In a sense, C acts as a natural experiment or manipulation of A, which, under the null hypothesis that A causes B, should result in a correspondence between C and B.

Scheines (2005) describes the similarities between causal inference in experimental and observational studies, and the conditions under which one can conclude that A does cause B even when A is not directly manipulated. As described below, two instrumental variables are required for such a conclusion, rather than just one—it is easier to disprove a causal relationship (as in BBH) using observational data than to prove one. These two instruments must be related to A and B, and to one another, in very specific ways. A causal inverted fork (involving A and the test instruments) is again crucial to the inference of the direction of causal flow, and distinguishes the methods described here from standard instrumental variables methods. If suitable instruments are available, then under some reasonable assumptions positive causal relationships can be formally inferred.

The present study seeks to provide evidence on the conviction with which one can conclude that A does cause B, where A, B, and related instruments are only observed and not actively manipulated. Couched in the philosophy of falsification and the framework of classical statistical hypothesis testing, we consider the conditions under which we can reject a null hypothesis that A does not cause B, which we write as H_0: $A \not\rightarrow B$, and the confidence that we can place on such judgements. That is, we conduct an analysis like that of BBH for the causal inference problem discussed in Scheines (2005).

2 Conditions that require rejection of H_0: $A \not\rightarrow B$

Two fundamental assumptions underlie causal inference using observational data. First, it is assumed (A1) that graphs representing a causal structure satisfy the Causal Markov Axiom, whereby all variables X are statistically independent of variables that are not effects of X, conditional on the direct causes of X. Second, it is assumed (A2) that joint probability distributions for variables in a graph are faithful to the underlying graph; that is they follow from the

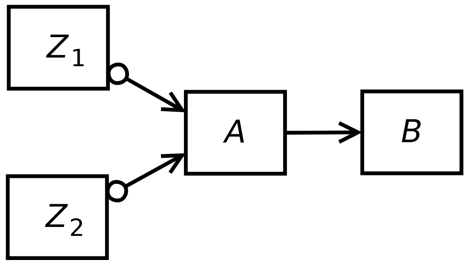

Figure 1: Graph representing observationally equivalent causal structures in which A must cause B

Causal Markov Axiom rather than peculiar values for the underlying structural parameters.[1] For example, we will not find two variables that are unconditionally statistically independent, despite being adjacent in the underlying causal graph.

The formal development of the logic underlying causal inference in the presence of latent variables is developed in Spirtes et al. (1999), and a subset of their fast causal inference (FCI) algorithm that is relevant to us here is discussed in Scheines (2005). Employing procedures that are appropriate for causal inference in the presence of unobserved, but causally related, variables frees the applied researcher from the onerous task of assembling observations of potentially large numbers of random variables. We consider four variables, Z_1, Z_2, A, and B, whose causal relationships are consistent with Figure 1. This figure is an example of a partial ancestral graph (PAG), which represents a class of possible causal graphs with identical independence relations among the observed variables. In a PAG, an edge such as $Z_1 \circ\!\!\rightarrow A$ indicates that either Z_1 causes A, or they share a latent common cause, or both. Note that the edge between A and B in Figure 1 embodies our alternative hypothesis, namely H_A: $A \to B$.

Specific independence relations are implied by this causal structure. Here, we are concerned with the relevant set of independence relations that implies (under A1 and A2, and the logic of the FCI algorithm) causal flow from A to B. We can reject H_0: $A \not\to B$ in favor of H_A: $A \to B$ if, and only if, all of the following conditions are satisfied:

1. $A \not\perp B$
2. $A \not\perp Z_1$
3. $A \not\perp Z_2$
4. $B \perp Z_1 | A$
5. $B \perp Z_2 | A$

[1]Spirtes et al. (2000) call this the Faithfulness assumption; Pearl (2000) calls this the Stability assumption.

6. $Z_1 \perp Z_2$

Here, $\perp$ denotes independence, $\not\perp$ denotes non-independence, and $|A$ indicates that independence is conditioned on the variable A.

In this arrangement, the observed variables Z_1 and Z_2 are providing natural manipulation of A, or evidence of natural manipulation of A by latent variables. Conditions two, three, and six convince us that we must have a subgraph in the equivalence class $Z_1 {\circ\!\!\rightarrow} A {\leftarrow\!\!\circ} Z_2$. If causal flow ran away from A towards either Z_1 or Z_2, then (given conditions two and three) condition six would not hold. Given manipulation of A by the instrumental variables Z_1 and Z_2 (or evidence of manipulation), then different independence relationships among B and the instrumental variables will follow under the null and alternative hypotheses.

Suppose the null hypothesis is true. Ignoring the degenerate case in which A and B are causally unrelated (implying that condition one is not satisfied), we therefore have $A {\leftarrow\!\!\circ} B$. Then we would have a partial causal structure consistent with $Z_1 {\circ\!\!\rightarrow} A {\leftarrow\!\!\circ} B$. Here, we would have $B \not\perp Z_1|A$ rather than $B \perp Z_1|A$. As an intuitive example, suppose A indicates if a sidewalk is wet or not, Z_1 indicates if sprinklers were on recently, and B indicates if it rained recently.[2] Then suppose we know that it did not rain recently (we know B), yet we observe a wet sidewalk (we also know A). Then we would conclude that the sprinklers were on recently (we can infer the likely state of Z_1). That is, $B \not\perp Z_1|A$.

Note that two instrumental variables are needed because having but a single instrument would potentially result in being unable to distinguish the direction of causal flow between A and the other variables. The independence relations implied by the causal structure $Z_1 \rightarrow A \rightarrow B$ are identical to independence relations for $Z_1 \leftarrow A \leftarrow B$. It is only the observation of two unconditionally independent (with one another) instrumental variables that allows the inference that causal flow runs from the instruments towards A, which facilitates the inference of the direction of causal flow between A and B using conditions four and five as we have just described.

The astute reader will notice that our conditions 1 through 5 correspond to conditions required by a standard instrumental variables (IV) analysis, although the specific statistical tests carried out would generally differ. Condition 1 ensures that there is an associate between A and B, conditions 2 and 3 ensure instrument relevance, and conditions 4 and 5 ensure instrument exogeneity. The primary purpose of IV estimation, however, is the consistent estimation of the marginal effect on B of a unit change in A, where the direction of causal flow (either from A to B or from the instrument(s) to A) has been assumed *a priori* (generally with theoretical justification). While a non-zero association between A on B discovered through IV estimation is sometimes interpreted as *proving* A causes B, this is not possible in the absence of an assumption (often implicit) regarding the direction of causal flow somewhere in the system. We demonstrate in Appendix A that all requirements under IV methods can be satisfied (including finding a non-zero association between A and B) even when B causes A. In the absence any assumptions regarding the direction of causal

[2]We adapt this example from Pearl (2000).

flow anywhere in the system, the formal causal inference literature and FCI algorithm unambiguously insist on condition 6 to orient the edge between A and B, as described above (Spirtes et al., 1999; Scheines, 2005). The requirement for two unconditionally independent instruments (our condition 6) distinguishes the method presented here from the standard IV approach.

We also note that one cannot prove that A causes B by artificially satisfying condition 6 using two naturally dependent instruments. One may be tempted, given two correlated instruments Z_1 and Z_2, to form an artificial instrument Z_2^* that is the residuals from a regression of Z_2 on Z_1. We show in Appendix A that using this revised system $\{A, B, Z_1, Z_2^*\}$, the remaining five conditions may be satisfied for systems in which B causes A. Thus, instruments that naturally satisfy condition 6 are required to reliably reject H_A: $A \to B$ in favor of H_A: $A \to B$.

3 Testing H_0: $A \not\to B$ for linearly related, jointly normal random variables

For the case of linearly related, jointly normal random variables, we can test H_0: $A \not\to B$ using Fisher's z-test to evaluate the independence relations described in Section 2. We adopt notation similar to that of BBH in this section. Assume that there is an underlying causal structure among m random variables which can be represented by a recursive structural equation model

$$X_n = \Gamma_0 + \Gamma_1 X_n + \epsilon_n \tag{1}$$

where X_n is a $m \times 1$ vector of covariates for observation n, including both the observed variables A, B, Z_1, and Z_2, and $m-4$ unobserved variables. ϵ_n is a conformable vector of independent, normally distributed errors. Γ_0 and Γ_1 are conformable parameter matrices, with non-zero elements of Γ_1 corresponding to edges present in a corresponding graph. For rows and columns of Γ_1 indexed by i and j, respectively, individual elements g_{ij} are zero for some ordering of the variables, reflecting the assumption of no cycles. Variables with non-zero elements in the ith row of $\Gamma_1 X_n$ cause the ith variable in X_n, but the reverse is not true.

For $\epsilon_n \sim N(0, \Sigma)$, where Σ is diagonal and $\Gamma_0 = 0$ without loss of generality, we have $X \sim N(0, C\Sigma C')$ for $C \equiv (I - \Gamma_1)^{-1}$. The unconditional population correlation coefficient between variables i and j is

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \tag{2}$$

where σ_{ij} is element i, j of $C\Sigma C'$. The population partial correlation coefficient between variables i and j, conditioned on a single variable k, is

$$\rho_{ij|k} = \frac{\rho_{ij} - \rho_{iz}\rho_{jz}}{\sqrt{1-\rho_{iz}^2}\sqrt{1-\rho_{jz}^2}} \tag{3}$$

For a sample $X_1, ..., X_N$, where x_{in} is the ith component of X_n, the corresponding sample correlation coefficients are

$$r_{ij} = \frac{\sum_{n=1}^{N} x_{in} x_{jn}}{\sqrt{\sum_{n=1}^{N} x_{in}^2} \sqrt{\sum_{n=1}^{N} x_{jn}^2}} \tag{4}$$

and

$$r_{ij|k} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{1 - r_{ik}^2} \sqrt{1 - r_{jk}^2}} \tag{5}$$

respectively.

We can test the null hypotheses H_0: $\rho_{ij} = 0$ and H_0: $\rho_{ij|k} = 0$ using the test statistics

$$z_{ij} = \frac{1}{2} \sqrt{N-3} \ln \left(\frac{|1 + r_{ij}|}{|1 - r_{ij}|} \right) \tag{6}$$

and

$$z_{ij|k} = \frac{1}{2} \sqrt{N-4} \ln \left(\frac{|1 + r_{ij|k}|}{|1 - r_{ij|k}|} \right) \tag{7}$$

respectively, where again k reflects only a single variable upon which we condition. Both z_{ij} and $z_{ij|k}$ are, to a close approximation, distributed as standard normal for small samples (Anderson, 2003).

Given our assumptions A1 and A2, $A \perp B \iff \rho_{AB} = 0$, $A \not\perp B \iff \rho_{AB} \neq 0$ and $B \perp Z_1 | A \iff \rho_{BZ_1|A} = 0$ and so forth. This equivalence means that we can test our overall hypothesis of H_0: $A \not\to B$ by conducting six Fisher's z-tests corresponding to the six independence relations given in the previous section. For this normal, linear case, the null hypothesis can be rejected if and only if all of the following conditions hold.

1. $\rho_{AB} \neq 0$
2. $\rho_{AZ_1} \neq 0$
3. $\rho_{AZ_2} \neq 0$
4. $\rho_{BZ_1|A} = 0$
5. $\rho_{BZ_2|A} = 0$
6. $\rho_{Z_1 Z_2} = 0$

Note that for conditions four through six, the burden of proof in the individual z-tests runs counter to that which we would desire for our overall hypothesis test: rejecting H_0: $A \not\to B$ relies upon us failing to reject H_0: $\rho_{BZ_1|A} = 0$, H_0: $\rho_{BZ_2|A} = 0$ and H_0: $\rho_{Z_1 Z_2|A} = 0$. This implies that the size and power of the overall test of H_0: $A \not\to B$ will not necessarily change monotonically as sample size increases, as is the case with conventional statistical tests. Size and power of the overall test will also interact in complex ways with the level of confidence employed in the underlying z-tests.

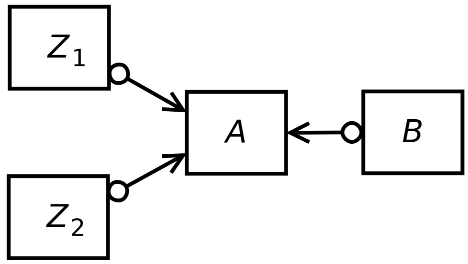

Figure 2: Graph representing equivalence class of causal structures under conditions of deliberate investigation of a true H_0: $A \not\to B$.

4 Monte Carlo simulations

To evaluate the empirical size and power of the the overall test of H_0: $A \not\to B$, we generate large numbers of random systems in which the null hypothesis is true and false, respectively, and observe the frequency with which the hypothesis is rejected. In the following two subsections, we individually describe for size and power simulations the generation of the random data sets, simulation procedures, and results.

4.1 Size

For size simulations, we generate a large number random data sets using randomly drawn causal structures in which the null hypothesis is true under likely conditions associated with deliberate investigation of a specific causal hypothesis.[3] Although there exists an infinite number of possible causal structures among our four observed variables and an infinite number of unobserved variables, many of these causal structures are uninteresting because they would not be encountered by an informed, deliberate investigator. The degenerate case where $A \perp B$ is unlikely to attract the attention of a researcher, and we do not consider this portion of the space of all potential causal structures. We further assume that the expert researcher will select IVs that are appropriate for the overall test. We therefore ignore the portions of the model space in which any of the following conditions are true: the IVs are directly causally related to B, the IVs are unrelated to A, or the IVs are unconditionally correspond with one another.

Given these constraints, we draw random causal structures from the equivalence class portrayed in Figure 2. Each edge depicted in Figure 2 has three possible states. Between the variables A and B for example, B can cause A in the absence of a latent common cause, B can directly cause A with a latent common cause also being present, or B and A are causally related only by a latent common cause. This implies $3^3 = 27$ possible causal structures from which

[3]Note that the results presented here are not applicable for inference in exploratory, algorithmic investigations of causal structures among large numbers of random variables.

we draw. We denote a latent common cause of A and B as L_{AB}, and common causes of the other relevant variable pairs similarly.

Each simulation begins with the selection of parameters d_0 and d_1 that will govern the random generation of structural parameters in equation 1. Without loss of generality, we assume Γ_0 is a zero vector for all systems. Each of the relevant elements of Γ_1 for a given causal structure is set at a level to corresponding to a correlation coefficient that is randomly drawn from a $U(d_0, d_1)$ distribution.[4] Note that X_n contains values not only for the observed variables A, B, Z_1, and Z_2, but also for unobserved variables L_{AB}, L_{AZ_1}, and L_{AZ_2}. Accordingly, some elements of Γ_1 reflect causal flow from the latent variables to the observed variables. Following Demiralp and Hoover (2003) and Bryant et al. (2009), we consider three expected signal strengths for randomly generated systems. We set the duple (d_0, d_1) to (0.01, 0.49), (0.25, 0.75), and (0.51, 0.99) to reflect low, medium, and high signal strengths in the simulations that follow.

For each signal strength specification, for each significance level α in the underlying z-tests in the set $\{0.05, 0.10, 0.15, 0.20, 0.25\}$, and for numbers of observations N (in the randomly generated data sets) in the set $\{50, 150, \ldots, 1000\}$, we conduct 10,000 trials. An individual trial consists of the following steps.

1. A random causal structure is drawn from the 27 causal structures in the equivalence class depicted in Figure 2
2. Γ_1 in equation (1) is randomly parametrized as described above, using the specified (d_0, d_1)
3. A random data set containing N observations is generated. For each observation n
   (a) The seven elements of ϵ_n are drawn independently from a standard normal distribution
   (b) Equation 1 is solved for $X'_n = [A, B, Z_1, Z_2, L_{AB}, L_{AZ_1}, L_{AZ_2}]_n$
4. The test of H_0: $A \not\rightarrow B$ described in Section 3 is applied using the N values for the observed variables A, B, Z_1, and Z_2 generated in step 3
5. The results of the test are recorded

Results of the size simulations are presented in Table 1, and graphically in Figure 3. Several aspects of these results are interesting. The results suggest that the overall test has reasonable size properties across all expected signal strengths, all considered values for α, and all considered sample sizes, with type I errors occurring in less than 15% of trials in call cases. Perhaps unsurprisingly, fewer

[4] The assigned structural coefficients correspond to the randomly drawn correlation coefficients supposing that one variable is the sole cause, in sense of equation 1, of the other variable. Thus the structural coefficients g_{ij} are given by $g_{ij} = \left(\rho_{ij}^{-2} - 1\right)^{1/2}$, where ρ_{ij} is generated randomly.

type I errors generally occur as signal strength increases. For a strong signal strength, the proportion of type I errors falls as the number of observations rises. This is not always the case for the medium signal strength, however, where for lower values of α the proportion of rejections initially rises as N increases. As discussed in section 3, this is due to the fact that the burden of proof in half of the six individual z-tests runs counter to that desired for the overall test of H_0: $A \not\rightarrow B$. This effect is even more pronounced for the weak signal strength cases, wherein size tends to rise for all values of α until about 700 observations are employed.

For the applied researcher with numerous observations, say 150 or more, who conservatively assumes that her data embody a weak expected signal strength, these size results suggest the specification of a relatively large value for α in the individual z-tests. This facilitates probabilities of type I errors in the 0.10 neighborhood, rather than the 0.15 neighborhood. This decision may be influenced however, by consideration of the influence of α on the expected power of the test.

4.2 Power

We conduct two types of power simulations. In both cases, the steps involved are identical to those for the size simulations, with the exception of the sets of causal structures from which we draw. Obviously, we now simulate data using causal structures in which H_0 is false. Similar to the size simulations, both types of power simulations assume conditions of deliberate and skilled investigation of H_0: $A \not\rightarrow B$. We therefore again do not employ many possible causal structures in which H_0 is false, but instruments are poorly selected.

The first set of power simulations assumes ideal conditions for inference. We assume that appropriate instruments are selected, that there is in fact an association between A and B, and that there are no latent variables present. Simulated systems are therefore based on the single causal structure shown in Figure 4. Simulations are again conducted for three expected signal strengths and various sample sizes and nominal sizes for underlying z-test . The proportions of trials in which we correctly reject H_0 are shown in Table 2, and graphically in Figure 5.

These results reveal that we are unlikely to reject a false null with small numbers of observations and a low signal strength, but improvements in either of these dimensions rapidly improves power. In almost all situations, specification of lower levels for the nominal size α in the underlying z-tests substantially improves power in the overall test of H_0: $A \not\rightarrow B$. This unfortunately is at odds with the size-based considerations for the specification of α that we presented in the previous subsection. However we have yet to consider the effects on power that might be caused by latent variables.

The second set of power simulations again assumes that instruments are well-selected, and that there is in fact a correspondence between A and B, but we now admit the possibility of latent variables. In these simulations, we draw from a set of causal structures that includes the nine that are consistent with

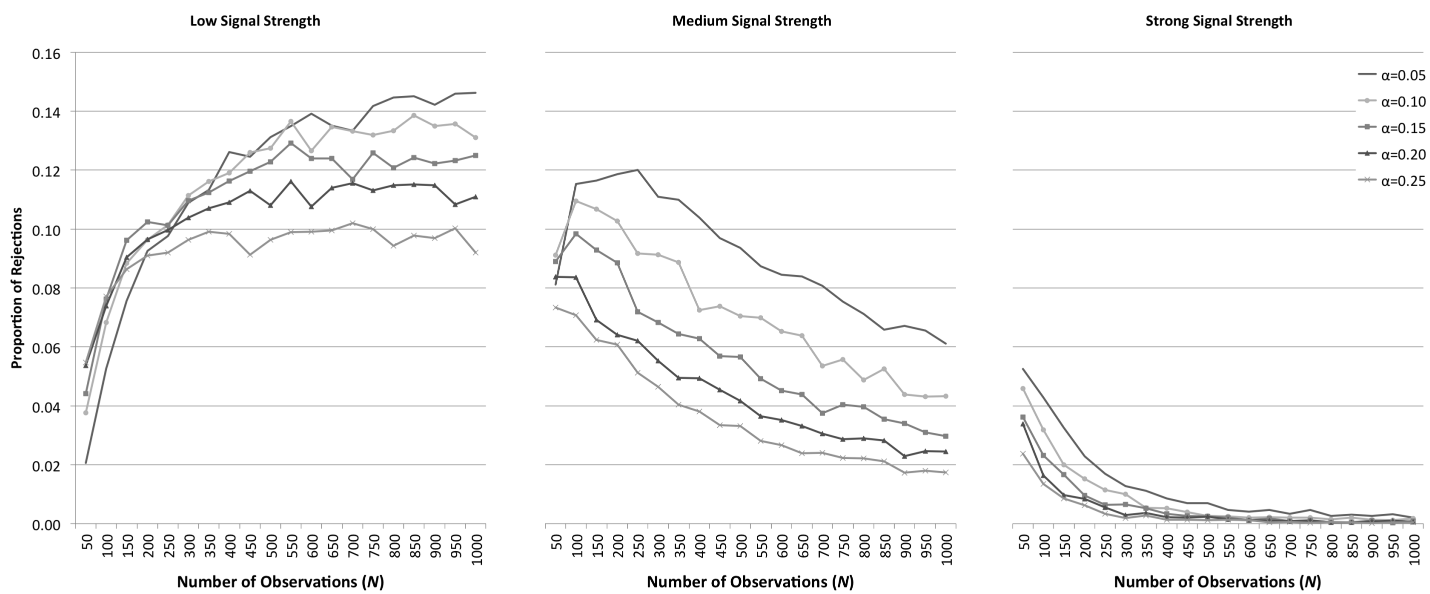

Figure 3: Size simulations: proportions of rejections when H_0: $A \not\to B$ is true

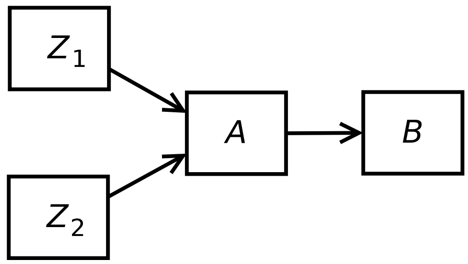

Figure 4: Causal structure in which H_0: $A \not\rightarrow B$ is false and no latent variables are present.

the PAG in Figure 1, and nine that are consistent with Figure 6. We draw from these 18 causal structures with equal probability. Results are presented in Table 3, and graphically in Figure 7.

A pattern similar to that of the ideal conditions power simulations is seen, albeit at lower levels overall. For many configurations of simulation parameters, the probability of rejecting a false H_0 goes down by 0.20 or more relative to the no latent variable simulations, indicating the advisability of selecting instruments that are directly causally related to A to the extent possible. We again note that for all but the lowest numbers of observations, power is increased, often substantially, by lowering the nominal size employed in the underlying z-tests.

Overall, for most numbers of observations, a substantial increase in power can be achieved by specifying a lower level of α at the expense of a modest deterioration of size in the overall test. This is portrayed graphically for the low signal strength case (for selected numbers of observations) in Figure 8. We therefore recommend that a low nominal size of $\alpha = 0.05$ or 0.10 be employed in the underlying z-tests, unless a researcher is extremely keen to avoid a type I error in the overall test.

5 Application

Gibson and Mace (2006) report significant increases in both the birth rate and childhood malnutrition in the Arsi area of Southern Ethiopia following improvements to the water infrastructure. They theorize the following causal flow underlies these associations: improved water access greatly reduced women's energetic expenditure on water collection and transport, resulting in increased fertility and greater scarcity of resources within households, ultimately resulting in increased childhood malnutrition. In this application, we seek to formally test the second half of this causal mechanism, namely that an increase in the birth rate (by whatever means) will cause an increase in malnutrition. Correct understanding of the causal mechanisms involved will facilitate improved policy. Gibson and Mace argue, based on their results, that to avoid increases in childhood

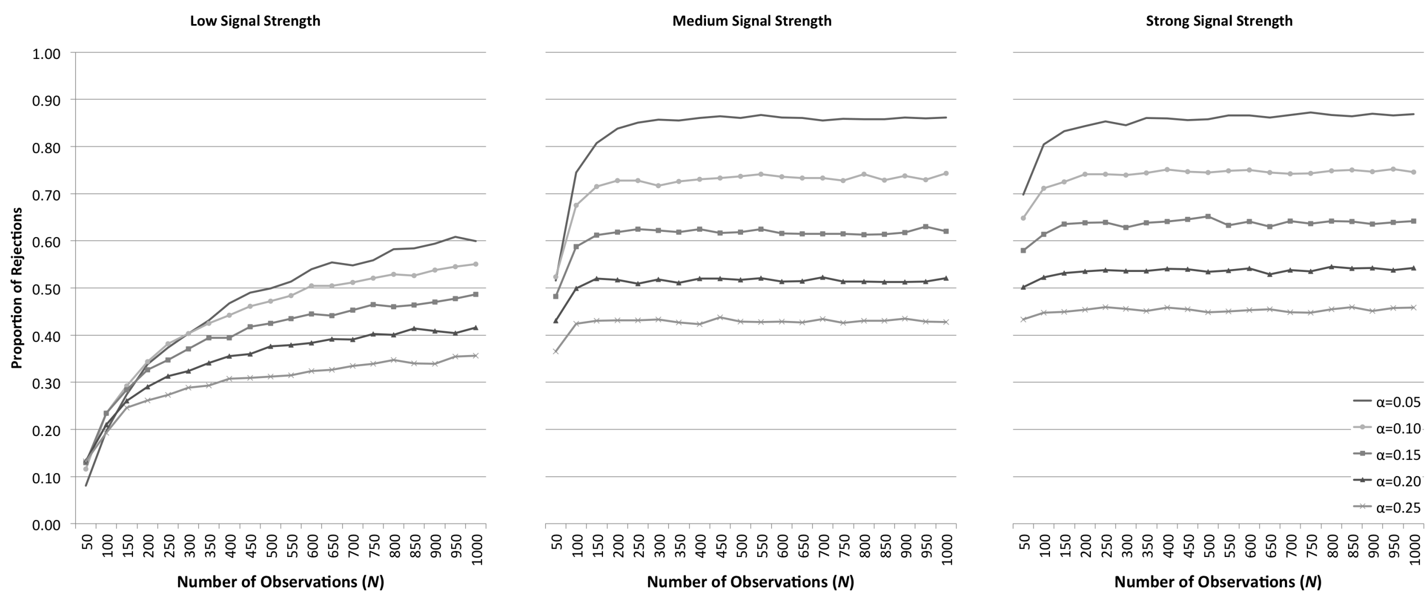

Figure 5: Power simulations: proportion of rejections when H_0: $A \not\to B$ is false, assuming no latent variables are present

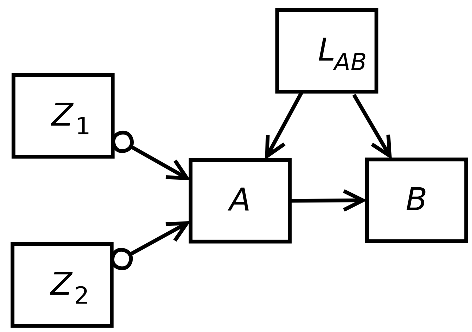

Figure 6: Class of causal structures in which H_0: $A \not\to B$ is false and one or more latent variables are present.

malnutrition, labor saving technological aid should include a family planning component. This policy recommendation is well-advised based on their hypothesized causal mechanism. If, however, the increase in childhood malnutrition that they document was not mediated by the increase in the birth rate, their policy recommendation would fail to prevent future increases in malnutrition.

We study late-1990's to mid-2000's cross section data on malnutrition, literacy, income distribution and birth rates from 110 countries. Data are from the World Bank Development Indicators and the CIA World Factbook. The countries studied are listed in Table B.1. Several "developed" countries are not on our list. This omission includes countries from Western Europe, North America and the Pacific-rim; the reason for exclusion is that the malnutrition numbers are not provided for many "developed" countries; e.g. United Kingdom, Canada, France, Japan, etc.

Malnourishment Prevalence (MAL), height for age, is reported in The World Bank Development Indicators tables for the 110 countries studied here. The two countries for which malnourishment data are reported, but that are not included in our analysis are Eritrea, for which we have no Gini coefficient, and Myanmar, for which we have no literacy rate. This measure is not available for each country on a continuous basis (malnutrition numbers are reported periodically for most countries). We used observations for the year 2006 or as close to 2006 as possible. These numbers appear to not vary greatly from year to year where continuous data are available. For example, the Romanian rate for 2000 was 13%, for 2001 it was 14%, and for 2002 it was 13%; nevertheless, we used the measure closest to 2006 which was 13%.

Birth Rate (BR). This measure is the number of live births occurring during the year 2006, per 1,000 population estimated at mid-year. The number is given for the year 2006 for each country in The World Bank Development Indicators tables.

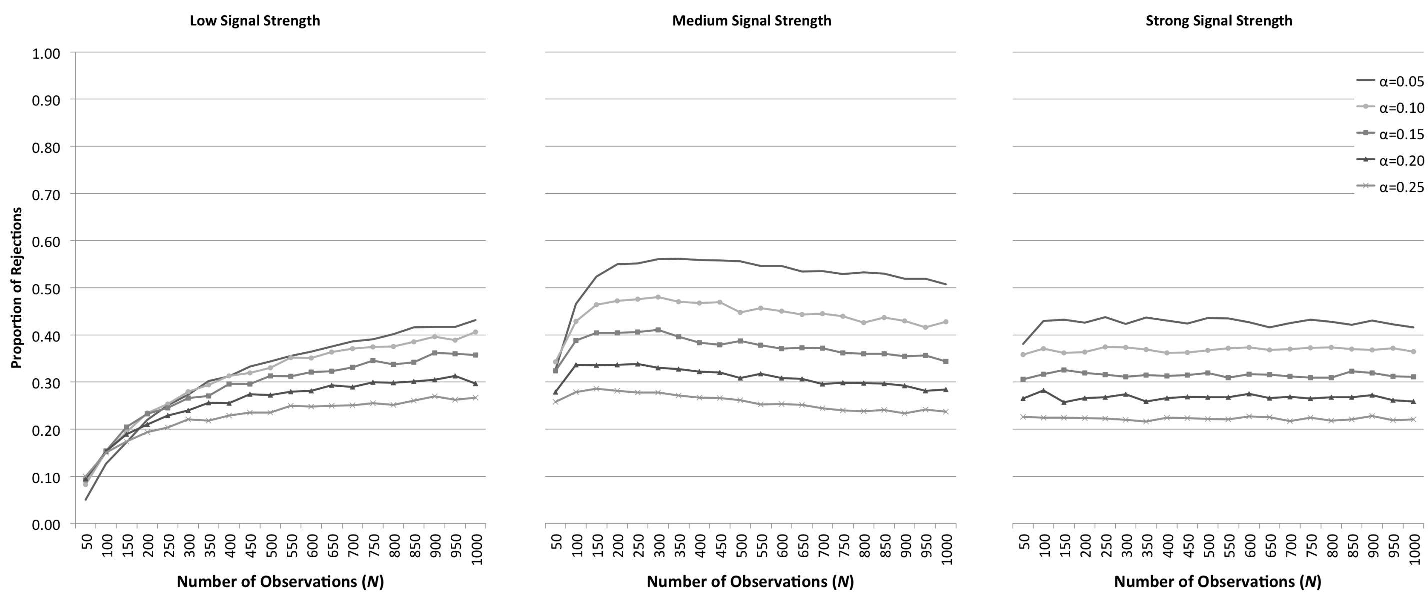

Figure 7: Power simulations: proportion of rejections when H_0: $A \not\to B$ is false, assuming latent variables are present

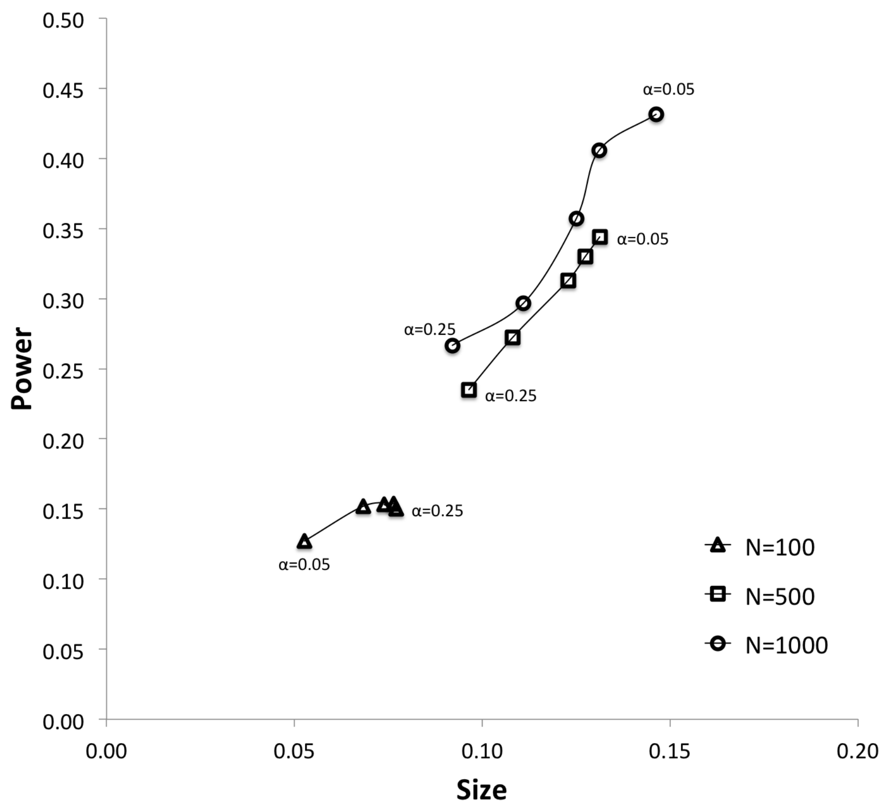

Figure 8: Size versus power for a test of H_0: $A \not\to B$, assuming a low signal strength

Literacy Rate (*LIT*). Here we use the numbers reported in the CIA World Fact Book tables as a measure of the percentage of people ages 15 and older that cannot, with understanding, read and write a short straightforward declaration on their daily life. Literacy rates are generally measured in the late-1990's or early-2000's.

Gini Index (*GINI*). This is a measure of income inequality, with zero indicating complete equality and 100 indicating complete inequality. These numbers are predominantly the UN measured Gini indices; where such numbers are not available we substituted the Global Peace Index Gini numbers, or if these numbers are not available we substituted CIA World Fact Book indices. Numbers used for each country are indicated in the note to Table B.1. Numbers are generally from mid-2000's measurements; although numbers on Algeria, Botswana, Burundi, Central African Republic, The Gambia, Guyana, Lesotho, Niger, and Sierra Leone pre-date 2000.

The test described in Section 3 and the FCI algorithm assume normally distributed data. From Table 4 we see that possible skewness exist for $GINI$ (positive) and LIT (negative). MAL and BR show some evidence of excess kurtosis. Accordingly, the results presented below should be viewed as an approximation on the underlying rejection probabilities.

We use these data to apply the test described in Section 3 to the hypothesis that birth rate does not cause malnutrition (H_0: $BR \not\rightarrow MAL$), with an alternative hypothesis that birth rate is a cause of malnutrition (H_A: $BR \rightarrow MAL$). Results of the individual underlying z-tests are presented in Table 5. Using the levels of significance in these underlying tests that we advocate in the previous section, we reject H_0: $BR \not\rightarrow MAL$, as all six conditions are satisfied. Referring to Table 1, we see that we expect the probability that this rejection reflects a type I error is ≤ 0.115 (the maximum entry across all three expected signal strengths in the 100 observations row and the $\alpha = 0.05$ and $\alpha = 0.10$ columns).

We can also describe the treatment of these data by the FCI algorithm to demonstrate a concordance with our test result. The lower triangular elements of the correlation matrix associated the data given in Table B.1 are the inputs into the FCI algorithm for calculation of correlation and partial correlations and associated test statistics to infer causal structure. This matrix is

$$\rho = \begin{array}{cc} \left[\begin{array}{cccc} 1.00 & & & \\ 0.20 & 1.00 & & \\ 0.36 & 0.78 & 1.00 & \\ -0.15 & -0.67 & -0.78 & 1.00 \end{array}\right] & \begin{array}{c} GINI \\ MAL \\ BR \\ LIT \end{array} \end{array} \tag{8}$$

Figure 9 gives the graphical pattern generated by the FCI algorithm using a p-value of 10% for underlying z-tests (as recommend by Spirtes, Glymour and Scheines 2000; the graphical structure at 5% is the same as that given here). Note that the PAG in this figure embodies our alternative hypothesis (H_A: $BR \rightarrow MAL$). The edge between LIT and $GINI$ is removed by zero

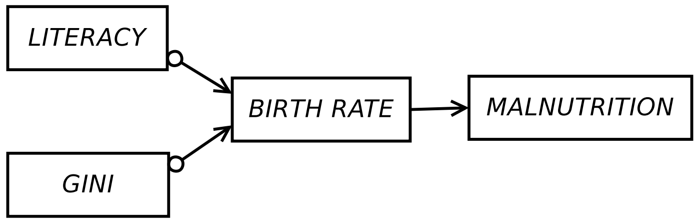

Figure 9: Graphical pattern on Gini index, literacy rate, birth rate, and malnutrition rate from 110 countries, late 1990's through mid 2000's data (described in Table B.1). This is the output of the FCI algorithm from the TETRAD project associated with the Philosophy Department at Carnegie Mellon University. A 10% significance level is used for edge removal as recommended in Spirtes et al. (2000). No prior "knowledge" (causal ordering) was used to find this graph structure with the FCI.

order conditioning, as the p-value on the correlation between $GINI$ and LIT is greater than 0.11, thus the null hypothesis that the partial correlation between $GINI$ and LIT is zero is not rejected. Similar hypotheses on all other zero-order correlations (between LIT and MAL, between LIT and BR, between $GINI$ and MAL, between $GINI$ and BR and between BR and MAL) are rejected. Additional higher-order conditioning results in failing to reject the null hypothesis of vanishing partial correlations (that the partial correlations equal zero) between MAL and LIT given BR (the partial correlation is -0.16, with a p-value of 0.10+), and the partial correlation between $GINI$ and MAL given BR (the partial correlation is -0.14 with a p-value of 0.15).

Accordingly, any information that emanates from LIT (or a latent which moves LIT) passes through BR in its flow to MAL. Similarly, $GINI$ information (or associated latents) flows through BR before reaching MAL. FCI cannot reject possible latents between LIT and BR and between $GINI$ and BR. But it can reject that a latent lies between BR and MAL. If a latent did in fact connect BR and MAL, the partial correlation between MAL and $GINI$ given BR or between LIT and MAL given BR would be non-zero (and they are not at p-values at or below 0.10).

Our formal causal analysis therefore supports the causal mechanisms hypothesized by Gibson and Mace. As an increased birth rate causes increased childhood malnutrition, policy makers should take care to avoid directly or indirectly increasing the birth rate with any planned aid intervention. Moreover, they could successfully reduce childhood malnutrition by actively reducing the birth rate.

6 Conclusions

We have described a means of rejecting a null hypothesis concerning observed, but not deliberately manipulated, variables of the form H_0: $A \not\rightarrow B$ in favor of an alternative hypothesis H_A: $A \rightarrow B$, even given the possibility of causally related unobserved variables. Rejection of such an H_0 relies upon the availability of two observed and appropriately related instrumental variables. The overall test is operational in the linear, jointly normal case using several underlying Fisher's z-tests of correlation among the observed variables, and we have characterized the confidence with which we can reject H_0: $A \not\rightarrow B$. This work is a natural complement to the that of BBH who characterized the conviction with which one might reject a null hypothesis of the form H_0: $A \rightarrow B$.

The test has attractive size properties. While the researcher will have limited control over the confidence level in this test, simulation results suggest that rejections of H_0: $A \not\rightarrow B$ occur at the 15% level or less. The power of the test is quite limited if there are but few observations available and the strength of correspondence among the variables is weak. The power of the test is generally improved by 1) larger numbers of observations, 2) improved correspondence between observed variables, and 3) specification of a fairly stringent threshold for rejection (α) in the underlying correlation tests. Changes in the power and size of the test are not monotonic as these factors change, however, because the the burdens of proof in the half of the underlying z-tests are opposite of that desired for the overall test. The above-mentioned factors therefore interact in complex ways to determine the size and power of the overall test. The careful selection of the best available instrumental variables Z_1 and Z_2 to employ, given expert knowledge of a particular problem domain, will be the primary challenge faced by an applied researcher.

We demonstrate the method by testing a hypothesis that the birth rate is a cause of childhood malnourishment using a cross section of data for 110 countries. We do not reject this hypothesis, with the implication that developmental aid interventions should be carefully designed to avoid directly or indirectly increasing birth rates.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Hoboken, New Jersey.

Bryant, H. L., Bessler, D. A., and Haigh, M. S. (2009). Disproving causal relationships using observational data. *Oxford Bulletin of Economics and Statistics*, 71(3):357–374.

Demiralp, S. and Hoover, K. D. (2003). Searching for the causal strucutre of a vector autoregression. *Oxford Bulletin of Economics and Statistics*, 70:509–533.

Gibson, M. A. and Mace, R. (2006). An energy-saving development initiative increases birth rate and childhood malnutrition in rural Ethiopia. *PLoS Med*, 3(4):e87.

Glymour, C. and Cooper, G. F., editors (1999). *Computation, Causation, and Discovery*. American Association for Artificial Intelligence, Menlo Park, California.

Pearl, J. (2000). *Causality*. Cambridge University Press, Cambridge, Massachusets.

Scheines, R. (2005). The similarity of causal inference in experimental and non-experimental studies. *Philosophy of Science*, 72:927–940.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, Massachusets.

Spirtes, P., Meek, C., and Richardson, T. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. In Glymour, C. and Cooper, G. F., editors, *Computation, Causality, and Discovery*. American Association for Artificial Intelligence, Menlo Park, California.

Table 1: Size simulations: proportion of rejections when H_0: $A \not\to B$ is true

	Low Signal Strength					Medium Signal Strength					Strong Signal Strength				
N	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$
50	0.021	0.038	0.044	0.054	0.055	0.081	0.091	0.089	0.084	0.073	0.053	0.046	0.036	0.034	0.024
100	0.052	0.068	0.076	0.074	0.077	0.115	0.110	0.098	0.084	0.071	0.043	0.032	0.023	0.016	0.014
150	0.076	0.088	0.096	0.090	0.086	0.117	0.107	0.093	0.069	0.062	0.033	0.020	0.012	0.010	0.009
200	0.093	0.096	0.102	0.097	0.091	0.119	0.103	0.089	0.064	0.061	0.023	0.015	0.010	0.008	0.006
250	0.098	0.101	0.101	0.100	0.092	0.120	0.092	0.072	0.062	0.051	0.017	0.011	0.006	0.006	0.003
300	0.109	0.111	0.110	0.104	0.096	0.111	0.091	0.068	0.055	0.047	0.013	0.010	0.007	0.003	0.002
350	0.113	0.116	0.112	0.107	0.099	0.110	0.089	0.064	0.050	0.040	0.011	0.005	0.005	0.004	0.003
400	0.126	0.119	0.116	0.109	0.098	0.104	0.073	0.063	0.049	0.038	0.009	0.005	0.003	0.002	0.001
450	0.125	0.126	0.120	0.113	0.091	0.097	0.074	0.057	0.045	0.033	0.007	0.004	0.003	0.002	0.001
500	0.131	0.127	0.123	0.108	0.096	0.094	0.070	0.057	0.042	0.033	0.007	0.003	0.003	0.002	0.001
550	0.135	0.137	0.129	0.116	0.099	0.087	0.070	0.049	0.037	0.028	0.005	0.003	0.002	0.001	0.001
600	0.139	0.127	0.124	0.108	0.099	0.085	0.065	0.045	0.035	0.027	0.004	0.002	0.001	0.001	0.001
650	0.135	0.135	0.124	0.114	0.100	0.084	0.064	0.044	0.033	0.024	0.005	0.002	0.002	0.001	0.001
700	0.133	0.133	0.117	0.116	0.102	0.081	0.054	0.038	0.031	0.024	0.003	0.002	0.001	0.001	0.001
750	0.149	0.132	0.126	0.113	0.100	0.075	0.056	0.040	0.029	0.022	0.005	0.002	0.001	0.001	0.000
800	0.145	0.133	0.121	0.115	0.094	0.071	0.049	0.040	0.029	0.022	0.003	0.002	0.001	0.001	0.001
850	0.145	0.139	0.124	0.115	0.098	0.066	0.053	0.035	0.028	0.021	0.003	0.002	0.000	0.001	0.000
900	0.142	0.135	0.122	0.115	0.097	0.067	0.044	0.034	0.023	0.017	0.003	0.001	0.001	0.001	0.000
950	0.146	0.136	0.123	0.108	0.100	0.066	0.043	0.031	0.025	0.018	0.003	0.001	0.000	0.001	0.001
1000	0.146	0.131	0.125	0.111	0.092	0.061	0.043	0.030	0.025	0.017	0.002	0.002	0.001	0.001	0.000

Table 2: Power simulations: proportion of rejections when H_0: $A \not\to B$ is false, assuming no latent variables are present

	Low Signal Strength					Medium Signal Strength					Strong Signal Strength				
N	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$
50	0.081	0.116	0.129	0.133	0.132	0.516	0.523	0.482	0.431	0.365	0.698	0.649	0.579	0.501	0.433
100	0.197	0.234	0.234	0.210	0.193	0.745	0.676	0.588	0.499	0.424	0.805	0.712	0.614	0.522	0.447
150	0.274	0.292	0.284	0.261	0.246	0.807	0.715	0.612	0.520	0.431	0.833	0.725	0.635	0.532	0.449
200	0.339	0.344	0.327	0.290	0.261	0.838	0.728	0.618	0.517	0.431	0.843	0.741	0.638	0.535	0.454
250	0.374	0.382	0.347	0.313	0.274	0.851	0.727	0.625	0.509	0.431	0.852	0.741	0.639	0.538	0.460
300	0.403	0.403	0.371	0.324	0.288	0.857	0.717	0.622	0.518	0.433	0.846	0.739	0.628	0.536	0.460
350	0.431	0.425	0.394	0.341	0.293	0.856	0.726	0.618	0.511	0.427	0.860	0.744	0.639	0.536	0.452
400	0.467	0.443	0.394	0.355	0.308	0.860	0.730	0.625	0.520	0.424	0.859	0.752	0.641	0.541	0.459
450	0.490	0.461	0.417	0.360	0.310	0.864	0.733	0.617	0.520	0.437	0.856	0.747	0.645	0.540	0.455
500	0.499	0.472	0.425	0.376	0.312	0.861	0.737	0.619	0.517	0.429	0.858	0.745	0.652	0.534	0.449
550	0.513	0.484	0.435	0.379	0.315	0.867	0.742	0.625	0.521	0.428	0.866	0.748	0.633	0.537	0.450
600	0.540	0.504	0.445	0.383	0.324	0.861	0.736	0.616	0.514	0.429	0.866	0.750	0.641	0.541	0.453
650	0.555	0.504	0.441	0.392	0.326	0.861	0.733	0.615	0.515	0.427	0.862	0.745	0.630	0.529	0.455
700	0.548	0.512	0.453	0.391	0.334	0.855	0.733	0.615	0.523	0.434	0.867	0.742	0.642	0.538	0.448
750	0.559	0.520	0.465	0.402	0.339	0.859	0.727	0.615	0.514	0.426	0.873	0.743	0.637	0.536	0.448
800	0.582	0.529	0.460	0.400	0.347	0.858	0.742	0.613	0.513	0.430	0.867	0.749	0.642	0.545	0.455
850	0.584	0.526	0.464	0.414	0.340	0.858	0.729	0.614	0.512	0.430	0.864	0.750	0.641	0.541	0.459
900	0.594	0.538	0.470	0.408	0.339	0.862	0.738	0.617	0.513	0.435	0.870	0.747	0.636	0.543	0.453
950	0.609	0.545	0.477	0.404	0.354	0.860	0.730	0.630	0.513	0.428	0.866	0.752	0.639	0.538	0.458
1000	0.599	0.551	0.486	0.416	0.357	0.862	0.743	0.620	0.521	0.427	0.869	0.745	0.641	0.542	0.458

Table 3: Power simulations: proportion of rejections when H_0: $A \not\to B$ is false, assuming latent variables are present

	Low Signal Strength					Medium Signal Strength					Strong Signal Strength				
N	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.15$	$\alpha = 0.20$	$\alpha = 0.25$
50	0.050	0.082	0.092	0.095	0.100	0.326	0.343	0.324	0.279	0.258	0.380	0.358	0.306	0.265	0.226
100	0.127	0.152	0.154	0.153	0.150	0.465	0.428	0.388	0.336	0.278	0.429	0.371	0.316	0.282	0.224
150	0.172	0.193	0.204	0.189	0.174	0.523	0.463	0.404	0.336	0.286	0.432	0.362	0.325	0.257	0.224
200	0.220	0.235	0.232	0.210	0.194	0.550	0.472	0.404	0.337	0.281	0.426	0.363	0.319	0.266	0.223
250	0.250	0.253	0.245	0.228	0.203	0.551	0.475	0.406	0.338	0.278	0.437	0.374	0.316	0.268	0.223
300	0.273	0.279	0.266	0.240	0.221	0.560	0.480	0.411	0.330	0.278	0.424	0.374	0.311	0.274	0.220
350	0.302	0.294	0.270	0.256	0.218	0.561	0.470	0.396	0.327	0.271	0.437	0.369	0.315	0.259	0.216
400	0.312	0.313	0.296	0.255	0.229	0.559	0.467	0.383	0.322	0.267	0.430	0.361	0.313	0.266	0.224
450	0.333	0.319	0.296	0.274	0.235	0.558	0.469	0.380	0.320	0.266	0.424	0.363	0.315	0.269	0.224
500	0.344	0.330	0.313	0.272	0.235	0.556	0.448	0.387	0.308	0.261	0.436	0.367	0.319	0.268	0.221
550	0.356	0.352	0.312	0.280	0.250	0.546	0.457	0.378	0.318	0.252	0.435	0.372	0.309	0.268	0.221
600	0.365	0.351	0.321	0.281	0.248	0.546	0.450	0.371	0.308	0.253	0.427	0.373	0.317	0.275	0.227
650	0.375	0.363	0.323	0.293	0.249	0.534	0.443	0.372	0.307	0.252	0.416	0.368	0.315	0.266	0.225
700	0.386	0.370	0.331	0.289	0.250	0.535	0.445	0.371	0.296	0.244	0.425	0.370	0.312	0.269	0.218
750	0.391	0.374	0.345	0.299	0.255	0.529	0.439	0.362	0.299	0.240	0.434	0.372	0.310	0.265	0.224
800	0.402	0.375	0.338	0.299	0.251	0.532	0.426	0.360	0.298	0.238	0.428	0.3731	0.309	0.2674	0.2181
850	0.416	0.386	0.342	0.301	0.260	0.530	0.437	0.360	0.297	0.241	0.421	0.3697	0.3227	0.268	0.2208
900	0.417	0.396	0.362	0.305	0.270	0.519	0.430	0.355	0.292	0.233	0.430	0.3678	0.3195	0.2718	0.2283
950	0.417	0.389	0.360	0.313	0.263	0.519	0.416	0.356	0.281	0.241	0.422	0.3715	0.3119	0.2614	0.2185
1000	0.431	0.406	0.357	0.297	0.267	0.508	0.428	0.344	0.284	0.237	0.416	0.3642	0.3111	0.2586	0.2209

Table 4: Descriptive statistics on measures of Gini index ($GINI$), malnutrition (MAL), birth rate (BR), and literacy rate (LIT) for late 1990's through mid 2000's data.

Variable	Mean (significance)	Skewness (significance)	Excess Kurtosis (significance)
$GINI$	43.49	0.434	-0.42
	(0.00)	(0.07)	(0.38)
MAL	29.44	0.06	-1.03
	(0.00)	(0.79)	(0.03)
BR	26.92	0.24	-0.96
	(0.00)	(0.30)	(0.05)
LIT	75.25	-0.71	-0.62
	(0.00)	(0.00)	(0.20)

Table 5: Test of H_0: birth rate (BR) $\not\rightarrow$ malnutrition (MAL), using Gini index ($GINI$) and literacy (LIT) as test instruments.

Hypothesis	Sample Correlation	z-statistic	p-value	Conclusion ($\alpha = 0.10$)
H_0: $\rho_{BR,MAL} \neq 0$	0.781	10.838	0.000	Reject
H_0: $\rho_{BR,GINI} \neq 0$	0.362	3.924	0.000	Reject
H_0: $\rho_{BR,LIT} \neq 0$	-0.776	-10.706	0.000	Reject
H_0: $\rho_{MAL,GINI\|BR} = 0$	-0.139	-1.445	0.148	Do not reject
H_0: $\rho_{MAL,LIT\|BR} = 0$	-0.157	-1.631	0.103	Do not reject
H_0: $\rho_{GINI,LIT} = 0$	-0.154	-1.607	0.108	Do not reject
H_0: $BR \not\rightarrow MAL$				Reject

Appendix A: Demonstrations of conditions that cannot prove causality

In this appendix, we demonstrate that conditions required by standard IV estimation may be satisfied in linear systems, including finding a significant coefficient in the second stage regression, when the regressor in the second stage regression is actually the cause of the regressand. This counter-example shows that satisfaction of the conditions required by standard IV estimation does not prove that the regressand in the second stage equation is the cause of the regressor.

To consistently estimate under the putative effect of A on B using instruments Z_1 and Z_2, a first stage regression of the form

$$A = \pi_1 Z_1 + \pi_2 Z_2 + v$$

is estimated, and the predicted values from this regression are employed in a second stage regression

$$B = \beta \hat{A} + u$$

The conditions required for a consistently estimated, non-zero "causal" effect of A on B (β) are judgements that

i. $\beta \neq 0$ (a non-zero "causal" effect)

ii. $\pi_1 \neq 0$ (Z_1 is a relevant instrument)

iii. $\pi_2 \neq 0$ (Z_2 is a relevant instrument)

iv. $E(Z_1 u) = 0$ (B and Z_1 are not correlated after conditioning on A)

v. $E(Z_2 u) = 0$ (B and Z_2 are not correlated after conditioning on A)

For linear systems, these conditions correspond to conditions 1 through 5 presented in Section 2. There is no condition for IV estimation corresponding to our condition 6.

Suppose that there is a recursive linear system where B is the sole cause of A, and A is the sole cause of both Z_1 and Z_2. Without loss of generality, we assume structural coefficients of zero for constants and unity for variable coefficients. We therefore have

$$B = \epsilon_B \tag{A.1}$$

$$A = B + \epsilon_A = \epsilon_A + \epsilon_B \tag{A.2}$$

$$Z_1 = A + \epsilon_{Z1} = \epsilon_A + \epsilon_B + \epsilon_{Z1} \tag{A.3}$$

$$Z_2 = A + \epsilon_{Z2} = \epsilon_A + \epsilon_B + \epsilon_{Z2} \tag{A.4}$$

with $E(\epsilon_i) = 0$ and $var(\epsilon_i) = \sigma_i^2 < \infty$ for $i \in \{A, B, Z_1, Z_2\}$, and $cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. The population first moments for the first stage coefficients are

$$\begin{bmatrix} \bar{\pi}_1 \\ \bar{\pi}_2 \end{bmatrix} = D \begin{bmatrix} \sigma_{Z2}^2 \\ \sigma_{Z1}^2 \end{bmatrix}$$

where

$$D = \frac{\sigma_A^2 + \sigma_B^2}{(\sigma_A^2 + \sigma_B^2 + \sigma_{Z1}^2)(\sigma_A^2 + \sigma_B^2 + \sigma_{Z2}^2) - 2(\sigma_A^2 + \sigma_B^2)}$$

These moments are greater than zero and conditions ii and iii are therefore satisfied. Using

$$\hat{A} = \bar{\pi}_1 Z_1 + \bar{\pi}_2 Z_2$$

the population first moment for the second stage coefficient is

$$\bar{\beta} = \frac{cov(\hat{A}, B)}{var(\hat{A})} = \frac{\sigma_B^2(\bar{\pi}_1 + \bar{\pi}_2)}{\bar{\pi}_1^2(\sigma_A^2 + \sigma_B^2 + \sigma_{Z1}^2) + \bar{\pi}_2^2(\sigma_A^2 + \sigma_B^2 + \sigma_{Z2}^2) + 2\bar{\pi}_1\bar{\pi}_2(\sigma_A^2 + \sigma_B^2)} > 0$$

and condition i is therefore satisfied. The second stage residuals are

$$\hat{u} = B - \hat{A}\bar{\beta}$$

and we can check conditions iv (and v analogously) by

$$\begin{aligned}
E(\hat{u}Z_1) &= E\left\{\epsilon_B Z_1 - \left(\bar{\pi}_1 Z_1^2 + \bar{\pi}_2 Z_1 Z_2\right)\bar{\beta}\right\} \\
&= \sigma_B^2 - \sigma_B^2\left[\frac{E\left(\bar{\pi}_1 Z_1^2 + \bar{\pi}_2 Z_1 Z_2\right)(\bar{\pi}_1 + \bar{\pi}_2)}{\bar{\pi}_1^2(\sigma_A^2 + \sigma_B^2 + \sigma_{Z1}^2) + \bar{\pi}_2^2(\sigma_A^2 + \sigma_B^2 + \sigma_{Z2}^2) + 2\bar{\pi}_1\bar{\pi}_2(\sigma_A^2 + \sigma_B^2)}\right] \\
&= \sigma_B^2 - \sigma_B^2\left[\frac{\left(\sigma_{Z2}^2\left(\sigma_A^2 + \sigma_B^2 + \sigma_{Z1}^2\right) + \sigma_{Z1}^2\left(\sigma_A^2 + \sigma_B^2\right)\right)\left(\sigma_{Z2}^2 + \sigma_{Z1}^2\right)}{\sigma_{Z2}^4(\sigma_A^2 + \sigma_B^2 + \sigma_{Z1}^2) + \sigma_{Z1}^4(\sigma_A^2 + \sigma_B^2 + \sigma_{Z2}^2) + 2\sigma_{Z1}^2\sigma_{Z2}^2(\sigma_A^2 + \sigma_B^2)}\right] \\
&= \sigma_B^2 - \sigma_B^2\left[\frac{\sigma_{Z2}^4\left(\sigma_A^2 + \sigma_B^2 + \sigma_{Z1}^2\right) + \sigma_{Z1}^4\left(\sigma_A^2 + \sigma_B^2 + \sigma_{Z2}^2\right) + 2\sigma_{Z1}^2\sigma_{Z2}^2\left(\sigma_A^2 + \sigma_B^2\right)}{\sigma_{Z2}^4(\sigma_A^2 + \sigma_B^2 + \sigma_{Z1}^2) + \sigma_{Z1}^4(\sigma_A^2 + \sigma_B^2 + \sigma_{Z2}^2) + 2\sigma_{Z1}^2\sigma_{Z2}^2(\sigma_A^2 + \sigma_B^2)}\right] \\
&= 0
\end{aligned}$$

All conditions for IV are satisfied, despite causality running from the second stage regressor to the regressand. Clearly, the second stage coefficient in a IV exercise can only be considered a causal effect of A on B given a correct assumption that any causality between A and B flows from the former to the latter, or that causality flows from the instruments to A.

We also demonstrate, using this same linear system (equations A.1 through A.4), that artificial satisfaction of our condition 6 using two naturally correlated instruments runs the risk rejecting a true H_0: $A \not\to B$. Suppose one forms an alternative instrument Z_2^* as the residuals from the regression of Z_2 on to Z_1:

$$\begin{aligned}
Z_2^* &= Z_2 - Z_1\hat{\gamma} \\
&= (1 - \hat{\gamma})\,\epsilon_A + (1 - \hat{\gamma})\,\epsilon_A + \epsilon_{Z1} - \hat{\gamma}\epsilon_{Z2}
\end{aligned}$$

where $\hat{\gamma}$ is the fitted coefficient. For the altered system $\{A, B, Z_1, Z_2^*\}$, condition 6 is artificially satisfied. Conditions 1, 2, and 3 are obviously satisfied for this system. Under linearity, condition 4 implies that the covariance between B and

Z_1, conditioned on A is zero. This is

$$
\begin{aligned}
E\left[\left(B-\mu_{B|A}\right)\left(Z_1-\mu_{Z_1|A}\right)\right] =& E\left\{BZ_1-\left[\frac{cov\left(A,B\right)}{var\left(A\right)}\right]AZ_1-\left[\frac{cov\left(A,Z_1\right)}{var\left(A\right)}\right]AB\right.\\
&\left.+\left[\frac{cov\left(A,B\right)cov\left(A,Z_1\right)}{var^2\left(A\right)}\right]A^2\right\}\\
=&\sigma_B^2-\left(\frac{\sigma_B^2}{\sigma_A^2+\sigma_B^2}\right)\left(\sigma_A^2+\sigma_B^2\right)-\left(\frac{\sigma_A^2+\sigma_B^2}{\sigma_A^2+\sigma_B^2}\right)\sigma_B^2\\
&+\left[\frac{\sigma_B^2\left(\sigma_A^2+\sigma_B^2\right)}{\left(\sigma_A^2+\sigma_B^2\right)^2}\right]\left(\sigma_A^2+\sigma_B^2\right)\\
=&\left(\sigma_B^2-\sigma_B^2-\sigma_B^2+\sigma_B^2\right)\\
=&0
\end{aligned}
$$

and therefore condition 4 is satisfied. Condition 5 is evaluated similarly.

$$
\begin{aligned}
E\left[\left(B-\mu_{B|A}\right)\left(Z_2^*-\mu_{Z_2^*|A}\right)\right] =& E\left\{BZ_2^*-\left[\frac{cov\left(A,B\right)}{var\left(A\right)}\right]AZ_2^*-\left[\frac{cov\left(A,Z_2^*\right)}{var\left(A\right)}\right]AB\right.\\
&\left.+\left[\frac{cov\left(A,B\right)cov\left(A,Z_2^*\right)}{var^2\left(A\right)}\right]A^2\right\}\\
=&\left(1-\hat{\gamma}\right)\sigma_B^2-\left(\frac{\sigma_B^2}{\sigma_A^2+\sigma_B^2}\right)\left(1-\hat{\gamma}\right)\left(\sigma_A^2+\sigma_B^2\right)\\
&-\left(\frac{\left(1-\hat{\gamma}\right)\left(\sigma_A^2+\sigma_B^2\right)}{\sigma_A^2+\sigma_B^2}\right)\sigma_B^2\\
&+\left[\frac{\sigma_B^2\left(1-\hat{\gamma}\right)\left(\sigma_A^2+\sigma_B^2\right)}{\left(\sigma_A^2+\sigma_B^2\right)^2}\right]\left(\sigma_A^2+\sigma_B^2\right)\\
=&\left(1-\hat{\gamma}\right)\left(\sigma_B^2-\sigma_B^2-\sigma_B^2+\sigma_B^2\right)\\
=&0
\end{aligned}
$$

Thus all six conditions are satisfied even though H_0: $A \not\to B$ is true for this system and H_A: $A \to B$ is false. Naturally independent instruments Z_1 and Z_2 are required to prove that A causes B.

Appendix B: Data used in example

Table B.1: Data and Sources

Country	$GINI$[5]	MAL[6]	BR[7]	LIT[8]
Afghanistan	60	59	48	28.1
Albania	31.1	27	14	98.7
Algeria	35.3	23	21	69.9
Angola	62	51	44	67.4
Argentina	51.3	8	17	97.2
Armenia	33.8	18	15	99.4
Azerbaijan	36.5	27	18	98.8
Bangladesh	33.4	43	23	47.9
Belarus	29.7	4	10	99.6
Belize	49.2	22	26	76.9
Benin	36.5	45	40	34.7
Bhutan	32	48	23	47
Bolivia	60.1	32	28	86.7
Bosnia and Herz	26.2	12	9	96.7
Botswana	60.5	29	25	81.2
Brazil	57	7	17	88.6
Bulgaria	29.2	9	10	98.2
Burkina Faso	39.5	44	47	21.8
Burundi	42.4	63	35	59.3
Cambodia	41.7	40	25	73.6
Cameroon	44.6	36	38	67.9
C. African Republic	61.3	45	37	48.6

[5] Gini Index data are taken from Wikipedia (http://en.wikipedia.org/wiki/List_of_countries_by_income_equality). All Gini numbers are UN measurements, except for numbers on: Afghanistan, Angola, Belize, Chad, Democratic Republic of Congo, Congo, Equatorial Guinea, Gabon, Guyana, Iraq, North Korea, Lebanon, Libya, Oman, Saudi Arabia, Somalia, Sudan, and Syria, which are Global Peace Index numbers. GINI observations for Guyana, Serbia, and Timor-Leste are from the CIA World Factbook. Accessed December 22, 2010.

[6] Malnutrition Prevalence, height for age, % of children under 5, are from the World Bank (http://ddp-ext.worldbank.org/ext/DDPQQ/showReport.do?method=showReport). Observations are for the closest available year to 2006. Accessed December 20, 2010.

[7] Birth Rate data are crude numbers of births per 1000 population for the year 2006 from the World Bank: (http://ddp-ext.worldbank.org/ext/DDPQQ/showReport.do?method=showReport). Accessed January 3, 2011.

[8] Literacy rates are from the CIA World Factbook: percent of population 15 and over who can read and write: (https://www.cia.gov/library/publications/the-world-factbook/fields/2103.html; Accessed December 27, 2010). The World Factbook numbers are similar, but not identical, to World Bank numbers measured more recently (but only sporadically) for several countries; e.g. our World Factbook numbers for India, Swaziland and Thailand are 61%, 81.6% and 92.6, respectively. World Bank numbers are 63% for India for the year 2006, 86% for Swaziland (2008) and 94% for Thailand (2005). As this latter data set is not measured for all countries in our study, we resort to the Factbook numbers.

Chad	52.3	45	47	25.7
Chile	54.9	2	15	95.7
China	46.9	22	12	91.6
Columbia	58.6	16	21	90.4
Dem Rep of Congo	55	46	46	67.2
Congo	56.2	31	35	83.8
Cote d'Ivoire	44.6	40	36	48.7
Czech Republic	25.4	3	10	99
Dominican Repub.	51.6	10	23	87
Ecuador	53.6	29	22	91
Egypt	34.4	31	25	71.4
El Salvador	52.4	25	20	81.1
Equatorial Guinea	65	35	38	87
Ethiopia	30	51	39	42.7
Gabon	60	26	28	63.2
The Gambia	50.2	28	38	40.1
Georgia	40.4	15	12	100
Germany	28.3	1	8	99
Ghana	40.8	28	33	57.9
Guatemala	55.1	54	34	69.1
Guinea	38.6	39	40	29.5
Guinea-Bissau	47	48	42	42.4
Guyana	43.2	18	19	91.8
Haiti	59.2	30	28	52.9
Honduras	53.8	30	28	80
India	36.8	48	24	61
Indonesia	34.3	40	19	90.4
Iraq	42	28	32	74.1
Jamaica	45.5	4	17	87.9
Jordan	38.8	12	26	89.9
Kazakhstan	33.9	18	20	99.5
Kenya	42.5	36	39	85.1
North Korea	31	45	14	99
Kyrgyzstan	30.3	18	23	98.7
Laos	34.6	48	28	73
Lebanon	45	16	16	87.4
Lesotho	63.2	45	30	84.8
Liberia	52.6	39	39	57.5
Libya	36	21	24	82.6
Macedonia	39	12	11	96.1
Madagascar	47.5	53	37	68.9
Malawi	39	53	41	62.7
Mali	40.1	38	43	46.4
Mauritania	39	29	34	51.2
Mexico	46.1	16	19	86.1
Moldova	33.2	11	12	99.1

Mongolia	32.8	28	19	97.8
Morocco	39.5	23	21	52.3
Mozambique	47.3	47	41	47.8
Namibia	74.3	30	28	85
Nepal	47.2	49	27	48.6
Nicaragua	43.1	19	25	67.5
Niger	50.5	55	54	28.7
Nigeria	43.7	43	41	68
Oman	32	13	22	81.4
Pakistan	30.6	42	31	49.9
Panama	56.1	22	21	91.9
Paraguay	58.4	18	25	94
Peru	52	30	22	92.9
Philippines	44.5	28	25	92.6
Romania	31	13	10	97.3
Rwanda	46.8	52	41	70.4
Saudi Arabia	32	9	24	78.8
Senegal	41.3	20	39	39.3
Serbia	26	8	10	96.4
Sierra Leon	62.9	47	41	35.1
Singapore	42.5	4	10	92.5
Somalia	30	42	45	37.8
Sri Lanka	40.2	17	19	90.7
Sudan	51	38	32	61.1
Swaziland	50.4	30	31	81.6
Syria	42	29	29	79.6
Tajikistan	32.6	33	28	99.5
Tanzania	34.6	44	42	69.4
Thailand	42	16	15	92.6
Timor-Leste	38	56	40	58.6
Trinidad & Tobago	38.9	5	15	98.6
Tunisia	39.8	9	17	74.3
Turkey	43.6	16	19	87.4
Uganda	45.7	39	47	66.8
Ukraine	28.1	23	10	99.4
United States	40.8	4	14	99
Uruguay	44.9	14	15	98
Uzbekistan	36.8	20	21	99.3
Vietnam	34.4	36	18	90.3
Yemen	33.4	58	37	50.2
Zambia	50.8	46	44	86.8
Zimbabwe	50.1	36	30	90.7