On Rejecting Causal Relationships with Observational Data

David A. Bessler
Texas A&M University

Herbert Simon Series – Seminar Number 8
Feng Chia University
Taichung, Taiwan
and
National Cheng-chi University
Taipei, Taiwan
R.O.C.

This presentation is based on joint work with Henry Bryant and Michael Haigh which is offered in the unpublished paper: Bryant, H., D. Bessler and M. Haigh, “Disproving Causal Relationships Using Observational Data,” Texas A&M University, December 2005.

Causal Relationships in Economics

Questions of causality have been central to economics from its beginnings, as the title of Adam Smith's *An Inquiry into the Nature and Causes of the Wealth of Nations* clearly indicates.

Economic theory is, at its most fundamental level, a body of hypotheses regarding causal relationships among economic variables – endowments, production, exchange, and consumption of goods, rates of exchange between goods and stores of value, aggregations of such quantities, and the evolution of these quantities over time.

Causality and Measurement in Economics

- Even though questions of causality are an integral part of economic theory, the practice of economic measurement has had an uneasy relationship with the matter.
- Haavelmo (1944) and other Cowles Commission econometricians devised structural equation models that explicitly represented hypothesized causal relationships.
- Explicit causal interpretations of these types of models have largely fallen out of favor, however, and they are today often interpreted simply as compact representations of joint probability distributions (Pearl, 2000).
- A perceived empirical failure of such structural modeling efforts motivated the extensive adoption of multivariate time series methods that have no clear causal interpretation (Heckman, 2000).

Present Day Practice has its Roots in the 19th Century

- Mill (1884) regarded causal inference using observational data as impossible, a sentiment that has been shared by many economists since.

- Even when confronted with empirical results that seem inconsistent with the causal content of economic theory, econometricians will generally assign blame to auxiliary hypotheses rather than questioning the theory itself (Blaug, 1992).

- Thus economic theories are generally “confirmed” or “verified”; rigorous testing of hypothesized causal relations in economics is sorely limited.

Recent Attempts to Ignore Mill's Impossibility

Despite Mill's beliefs on the matter, many scholars have begun to infer causal relationships from observational data.

Reichenbach (1956) proposed that causal relationships among random variables have specific implications for associated statistical independence relations.

Hausman (1983) is an early acknowledgment in economics that such causal inference should be feasible.

More recently, several algorithms for conducting such inference have been proposed (Spirtes, Glymour and Scheines, 2000; Glymour and Cooper, 1999; and Pearl, 1995, 2000).

Causal Inference Algorithms

The designers of these causal inference algorithms seem to intend them to be used in a manner that might be described as “data mining”, or “machine learning”.

In such use, observations of a large number of potentially related variables are assembled, and a causal structure among those variables is inferred. Most proposed algorithms conduct this overall inference by sequentially conducting several individual tests of conditional independence among the variables.

This multiple testing leads to criticism that the overall probability of an error is unknown, and possibly unreasonably high, particularly for a large system.

Monte Carlo Evidence

Casual experimentation with the algorithms using data sets with a moderate to high number of variables suggests that results are indeed fragile, and reversals of the direction of causal flow are not uncommon as one changes the algorithms' parameters.

Based on Monte Carlo experimentation, Spirtes, Glymour and Scheines (2000) suggest a negative relationship between appropriate p-values and number of observations.

Swanson and Granger (1997) and Dermilap and Hoover (2003) investigate such issues, finding that the probabilities of such errors are sensitive to the peculiarities of the data sets and difficult to quantify.

Today's Presentation

- We investigate the use of causal inference methods for testing a specific hypothesized causal relation:

  Ho: A causes B.

  We use notions developed from the machine learning literature to inform our confidence in rejection probabilities. The inference algorithms discussed above cast new light on hypothesis rejections.

- Our main finding is that these inference algorithms allow researchers to qualify hypothesis rejections as either weak basis rejections or strong basis rejections.

Preview of Results

A small number of causally-related variables are needed.

The researcher needs not observe all potentially causally-relevant variables. Observing some variable C that is causally-related in the right way to A and B should allow us to reject Ho, regardless of what other causally-related variables may exist.

Testing such a hypothesis with respect to a particular C involves only three individual tests of unconditional independence, allowing us to numerically estimate the size of such a test.

Causal Inference Algorithm

- We use ideas embedded in the Causal Inference Algorithm of Spirtes, Meek and Richardson (1999).

- While their algorithm works for arbitrary number of observed variables (V_i , $i = 1, \ldots, n > 3$), we explore testing causal hypotheses when $n = 3$.

- Latent or unobserved variables are formally introduced

Assumptions

- A1: Reichenbach's (1956) principle of the common cause holds: two variables are statistically dependent if and only if one variable causes the other or they share one or more common causes.

- A2: Faithfulness: two variables that share a common cause will not be rendered independent by off-setting parameter values. Independence relations reflect underlying causal structure and not off-setting parameter values (see Spirtes, Glymour and Scheines 2000 for more here).

Cases of Interest in Testing A does not cause B

Suppose A and B are not statistically independent we denote this by $A \not\perp B$. Then four cases are of interest:

- Case 1: $A \perp C$ and $B \not\perp C$. By A1 and A2 and the independence of A and C we must conclude: A cannot cause C either directly or indirectly. We must conclude: A o→ B ←o C.
- Case 2: $A \not\perp C$ and $B \perp C$. As with case 1 we must conclude: A cannot cause C either directly or indirectly. We must conclude: C o→ A ←o B.

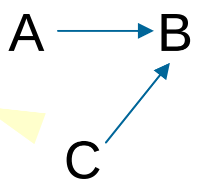
Cases of Interest, Continued

- Case 3: $A \perp C$ and $B \perp C$. Here C is not related to A and B and thus provides no information regarding the hypothesis a causality between A and B.

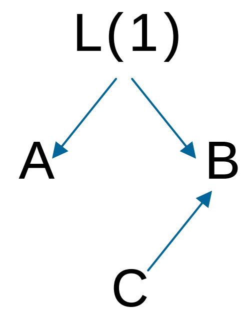
- Case 4: $A \perp C$ and $B \not\perp C$. Here there is no basis for rejecting the hypothesis $A \not\rightarrow B$.

Case I has three interesting possibilities

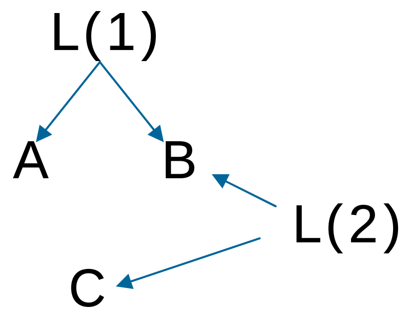
(i)



(ii)



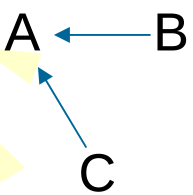
(iii)



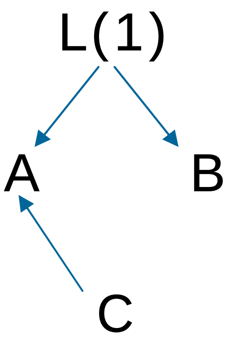
Any are possible under case I, thus there is no clear evidence that A does not cause B. Here L(1) and L(2) are latent variables.

Case II also has three interesting possibilities

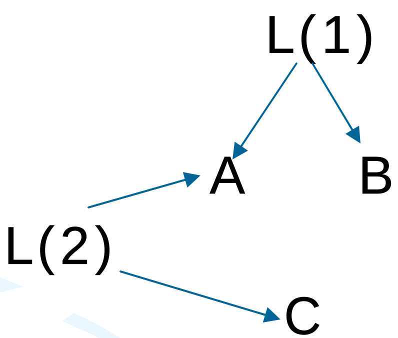
(i)



(ii)



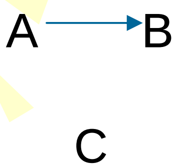
(iii)



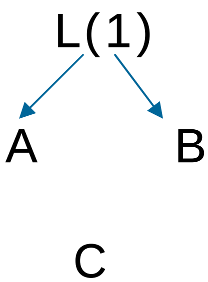
In all cases here there is evidence that A does not cause B.
Again L(1) and L(2) are latent variables.

Case III has three interesting possibilities

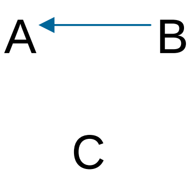
(i)



(ii)

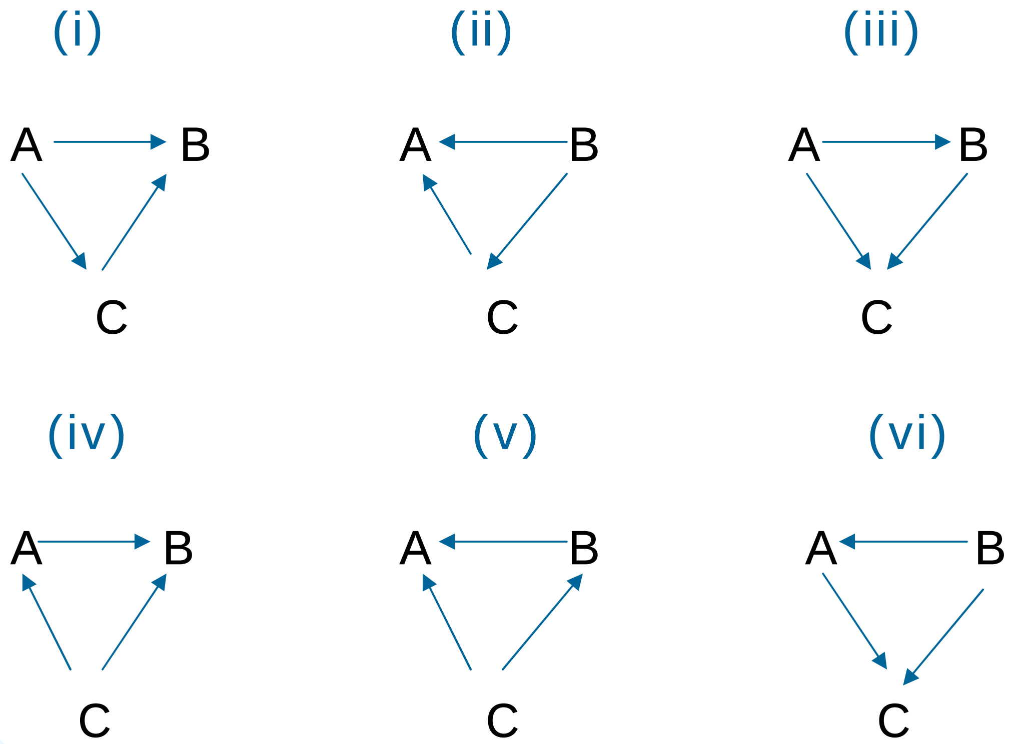


(iii)



Since any of these three are possible, case III provides no clear evidence that A does not cause B.

And Finally, Case IV has several possibilities



Or we could have latent variables mediating between each pair of variables A, B, and C. There is no clear evidence that A does not cause B.

Two Ways to Reject A Causes B

- From the above four cases and each possibility associated with each case we saw that it is only case II that gives us unambiguous evidence to reject the hypothesis that A causes B.

  [Recall in all four cases and their possibilities we observed a correlation between A and B.]

- The other case [not mentioned above] that allows us to reject the hypothesis that A causes B is if we observe a zero correlation between A and B.

- We refer to these two cases below as strong basis rejections and weak basis rejections, respectively.

Instrument for Testing A causes B

- Based on the above we say that C is an instrument for testing A causes B.

- To reject A causes B in “strong basis” form we need to find A correlated with B, A correlated with C and C not correlated with B.

- That is we need to find, in Pearl’s language, the following inverted fork between A, B and C:

$$C \rightarrow A \leftarrow B.$$

This requires $\rho(A,B)$⑤$0$; $\rho(A,C)$⑤$0$; and $\rho(C,B)=0$.

Tests of Correlations: Fisher's Z

Fisher's *z* statistic can be applied to test for significance from zero:

$$z(\rho(i,j);n) = 1/2(n-3)^{1/2} \times \ln\{(|1 + \rho(i,j)|) \times (|1 - \rho(i,j)|)^{-1} \} .$$

n is the number of observations used to estimate the correlations, $\rho(i,j)$ is the population correlation between variable i and j .

Monte Carlo Experiments

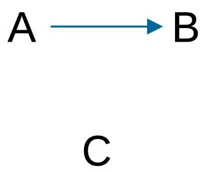
- Consider all possible directed acyclic graphs associated with variable A, B and C. There are twenty five such cases. Of these nine involve structures such that A causes B (either directly or indirectly through C). See Haigh and Bessler *Journal of Business* 2004 for details.

- We include as well latent variables between variable A, B and C, as L_{AB} L_{AC} and L_{BC} . Each of these may or may not be present in the system.

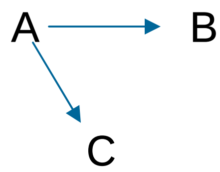
- Thus there are $2^3 = 8$ possible arrangements of latent variables that may accompany the 9 causal structures (of the 25 discussed above) in which A causes B. This gives us $9 \times 2^3 = 72$ causal structures.

DAGs where A causes B

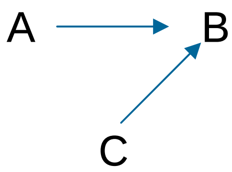
1



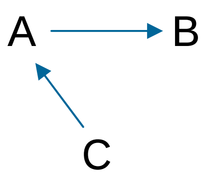
2



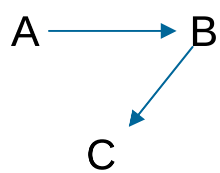
3



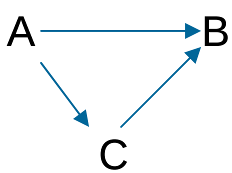
4



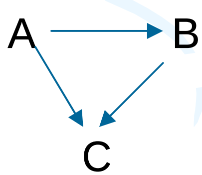
5



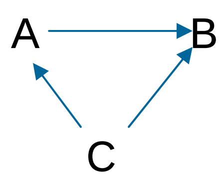
6



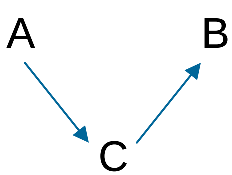
7



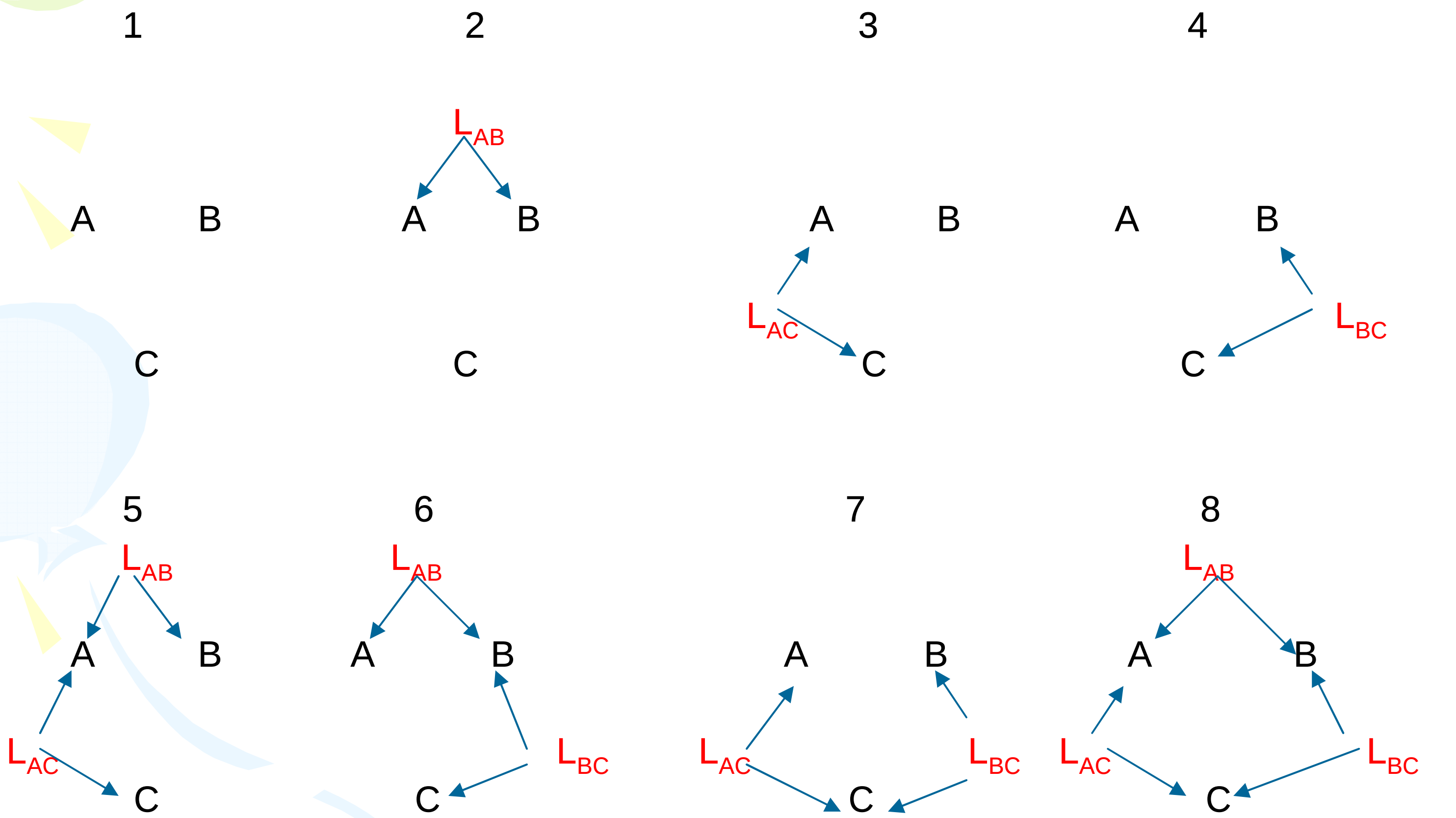
8



9



For each of the Nine DAGs on the Previous Slide the Following Eight Latent Structures are Considered



Linear Structural Representation

Each of the 72 DAGs can be represented as a recursive structural equation model:

$$(1)\ X = \Gamma_o + \Gamma_1 X + \varepsilon$$

Where Γ_o is a conformable intercept parameter matrix, Γ_1 is a lower triangular matrix reflecting the 72 DAGs discussed above, and ε is a conformable innovation vector with finite variance.

In our case here X' = (A, B, C, L_{AB} , L_{AC}, L_{BC}).

Draws

- We randomly select one such DAG, use that DAG to parameterize equation 1 (previous slide), where parameters of Γ_o are set to zero and Γ_1 are set to reflect three signal strengths following Demiralp and Hoover (2003).

- Individual elements (A,B,C) are drawn from a uniform mean zero variance d distribution. So that the mean parameter value will result in one of the three desired population correlation coefficients (weak signal, medium signal and strong signal) between the variables.

- ε is drawn from a N(0,1) distribution.

Monte Carlo Results I

Proportion of Rejections of a true Hypothesis A causes B - alpha is .10

	Low Signal	Medium Signal	High Signal
N=50			
Weak-basis	0.421	0.125	0.047
Strong-basis	0.093	0.044	0.012
N=250			
Weak-basis	0.174	0.043	0.017
Strong-basis	0.100	0.025	0.007

Monte Carlo Results II

Proportion of Rejections of a true Hypothesis A causes B - alpha is .20

	Low Signal	Medium Signal	High Signal
N=50			
Weak-basis	0.331	0.095	0.036
Strong-basis	0.110	0.043	0.013
N=250			
Weak-basis	0.1344	0.043	0.017
Strong-basis	0.100	0.025	0.007

Conclusions

The researcher need not observe all potentially causally-relevant variables. Observing some variable C that is causally-related in the right way to A and B should allow us to reject Ho, regardless of what other causally-related variables exist.

Testing such a hypothesis with respect to a particular C involves only three individual tests of unconditional independence, allowing us to numerically estimate the size of such a test.

Conclusions Continued

- Such testing gives us weak basis rejections and strong basis rejections.
- Weak basis rejections of the hypothesis follow from the sharp hypothesis failure to reject $\rho(A,B) = 0$.
- Strong basis rejections follow from rejection of the hypothesis $\rho(A,B) = 0$, $\rho(A,C) = 0$, and failure to reject $\rho(B,C) = 0$.
- Strong basis rejections are more reliable, reflecting their more strenuous testing conditions.

Conclusions Continued

- Rejecting a causal hypothesis is much easier than proving a causal hypothesis.

- Causal hypotheses not rejected with a particular instrument C must await further testing with a yet to be discovered instrument C*.

- We conclude with the asymmetry: rejected causal hypotheses based on the instrument C, remain rejected when the larger set C and C* is studied. Causal hypotheses not rejected under instrument C, may well be rejected under instrument C*.