The Role of Assumptions in Causal Discovery

Marek J. Druzdzel

Decision Systems Laboratory, School of Information Sciences and Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA 15260, USA

Faculty of Computer Science, Białystok Technical University, Wiejska 45A, 15-351 Białystok, Poland

marek@sis.pitt.edu

Abstract

The paper looks at the conditional independence search approach to causal discovery, proposed by Spirtes *et al.* and Pearl and Verma, from the point of view of the *mechanism-based* view of causality in econometrics, explicated by Simon. As demonstrated by Simon, the problem of determining the causal structure from data is severely underconstrained and the perceived causal structure depends on the *a priori* assumptions that one is willing to make. I discuss the assumptions made in the independence search-based causal discovery and their identifying strength.

1 Introduction

An accepted scientific procedure for demonstrating causal relations is experimentation. If experimental manipulation of one variable (called the independent variable) results in a change in value of another variable (called the dependent variable), assuming an effective control for all possible intervening variables, one usually concludes that in the system under study the two variables stand in a causal relation with each other. Unfortunately, conducting such experiments is for many practical systems impossible, because of our inability to manipulate the system variables, forbidding costs of experimentation, or ethical considerations. Numerous examples of such systems are found in economics, medicine, meteorology, or social sciences. Still, one wants to predict the impact of policy decisions, such as whether to impose a tax, introduce or abolish the death penalty, or restrict smoking, on such variables as the gross national product, crime rates, or the number of lung cancer cases in the population. Where experimentation is impossible, one must rely on observations and assumptions in order to form a theory of causal interactions.

One discipline where much attention has been paid to model construction from observations is econometrics. Work in late 1940s and early 1950s (see for example [4] or [3]) concentrated on formulating economic theories in the form of systems of structural equations, i.e., equations describing mechanisms by which variables interact directly with each other. It was commonly believed that systems of structural equations should be formulated either entirely on the basis of economic theory or economic theory combined with systematically collected statistical data for the relevant variables in the system. Construction of a system in the second case consisted of proposing a theoretical model, i.e., specifications of the form of the structural equations (including designation of the variables occurring in each of the equations) and then estimating the constant parameters from observations. The limits of such estimation raised the problem of "identifiability," i.e., whether it is theoretically possible, given prior knowledge about the functional forms of equations in a set of simultaneous equations, to determine unique values of parameters of these equations from observations. Simon [7] related the problem of identifiability to the causal structure of the system, showing theoretical conditions under which a structure is identifiable.

In their influential work, Spirtes *et al.* [9] and Pearl and Verma [5],[1] proposed that, under certain circumstances, observation is sufficient to determine all or part of the causal structure of a system. They have outlined methods for identifying the narrow class of causal structures (ideally a unique causal structure) that are compatible with particular observations. I will refer to the view of causality that underlies this work as *independence search-based* view of causality (or briefly ISC). As Simon [8] demonstrated, the problem of determining the causal structure from data is severely underconstrained and the perceived causal structure depends on the *a priori* assumptions that one is willing to make. From this point of view, there is little doubt that these new methods rest on some powerful identifying assumptions.

The goal of this paper is to explicate these assumptions, express them in terms of the earlier work in econometrics on structural equation models, and discuss their identifying strength. I will build on the results presented in [2], which reviews the mechanism-based view of causality (MBC) and shows a link between causal ordering and directed probabilistic graphs. The main conclusion resulting from this analysis is that with respect to the meaning of causality, the ISC and MBC views are almost identical. The power of the new methods rests on additional assumptions about causal relations that had not been made in econometrics. The two new powerful identifying assumptions are (1) that the causal structure is acyclic and (2) that each observed independence and dependence is a reflection of the causal structure and not merely coincidental (the latter called in the ISC view "faithfulness assumption"). With respect to the faithfulness assumption, the new, previously unexplored, element is dependence of causes conditional on a common effect.

The remainder of the paper is structured as follows. Section 2 starts with a

[1] I will refer frequently to the book by Spirtes *et al.* [9] rather than to the work of Pearl and Verma, because I am more familiar with the former. I believe that for the purpose of this analysis, both approaches are equivalent. There are other, Bayesian approaches to causal discovery originating from the seminal work of Cooper and Herskovitz [1], which I will leave outside this discussion.

brief review of the mechanism-based view of causality in directed probabilistic graphs. Section 3 offers a summary of the main assumptions made in the causal discovery work. Section 4 covers important concepts at the foundations of causal discovery: independence, conditioning, Markov condition, and faithfulness. It proposes a deterministic notion of independence and explains the link between this and the probabilistic view. Section 5 translates the assumptions in ISC into the MBC and explicates their identifying power.

2 Mechanism-Based View Of Causality

The mechanism-based view of causality rests on the observation that individual causal mechanisms, while normally symmetric (e.g., forces are reciprocal), exhibit asymmetry when embedded in the context of a model. Simon [7] proposed a procedure for deriving a directed graph of interactions among individual variables, called *causal ordering*, and tied it to the econometric notion of structure. He postulated that when each of the equations in the model is structural and each of the exogenous variables is truly exogenous, the asymmetry reflects the causal structure of the system. Druzdzel and Simon [2] have shown the link between causal ordering and directed probabilistic graphs. I will briefly review the main results from that work.

The following theorem demonstrates that the joint probability distribution over n variables of a Bayesian network (BN) can be represented by a model involving n simultaneous equations with these n variables and n additional independently distributed latent variables.

Theorem 1 (representability) *Let $\mathcal{B}$ be a BN model with discrete random variables. There exists a simultaneous equation model $\mathcal{S}$, involving all variables in $\mathcal{B}$, equivalent to $\mathcal{B}$ with respect to the joint probability distributions over its variables.*

The following theorem establishes an important property of a structural equation model of a system with the assumption of causal acyclicity.

Theorem 2 (acyclicity) *The acyclicity assumption in a causal graph corresponding to a self-contained system of equations $\mathcal{S}$ is equivalent to the following condition on $\mathcal{S}$: Each equation $e_i \in \mathcal{S} : f(x_1, \ldots, x_n, \mathcal{E}_i) = 0$ forms a self-contained system of some order k and degree one, and determines the value of some argument x_j $(1 \leq j \leq n)$ of f, while the remaining arguments of f are direct predecessors of x_j in causal ordering over $\mathcal{S}$.*

The last theorem binds causal ordering with the structure of a directed probabilistic graph.

Theorem 3 (causality in BNs) *A Bayesian belief network $\mathcal{B}$ reflects the causal structure of a system if and only if (1) each node of $\mathcal{B}$ and all its direct predecessors describe variables involved in a separate mechanism in the system, and (2) each node with no predecessors represents an exogenous variable.*

The above results show a link between structural equation models and causal graphs. They also make it clear that the former give a more general notion of structure than the letter. Directed probabilistic graphs are acyclic, while causal ordering in structural equation models can lead to cyclic structures. While equations can easily model dynamic processes with feedback loops, directed acyclic graphs can capture only their equilibrium states.

Theorem 3 demonstrates that directed arcs in BNs play a role that is similar in its representational power to the structure (presence or absence of variables in equations) of simultaneous equation models. The graphical structure of a BN, if given causal interpretation, is a qualitative specification of the mechanisms acting in a system.

3 Causal Discovery

Causal discovery in ISC is based on two axioms binding causality and probability. Informally, the first axiom, *causal Markov condition*, states that once we know all direct causes of an event, the event is probabilistically independent of its causal non-descendants. For example, suppose that we see a broken glass bottle on the bicycle path with small pieces of glass lying all around. Learning the cause of this broken bottle or that a piece from the bottle hurt a passing dog, does not change our expectation of a flat tire caused by the pieces of glass on the road.[2] The formal statement of the causal Markov condition is as follows:

> **Causal Markov Condition:** [9, page 54,] Let $\mathcal{G}$ be a causal graph with vertex set $\mathbf{V}$ and $\mathcal{P}$ be a probability distribution over the vertices in $\mathbf{V}$ generated by the causal structure represented by $\mathcal{G}$. $\mathcal{G}$ and $\mathcal{P}$ satisfy the Causal Markov Condition if and only if for every W in $\mathbf{V}$, W is independent of $\mathbf{V}\backslash(\text{Descendants}(W) \cup \text{Parents}(W))$ given $\text{Parents}(W)$.

The second axiom, the *faithfulness condition*, assumes that all interdependencies observed in the data are structural, resulting from the structure of the causal graph, and not accidental (e.g., by some particular combination of parameter values that result in causal effects canceling out). Spirtes *et al.* demonstrate that purely accidental dependencies and independences have, under a wide class of natural distributions over the parameters, a probability of measure zero. The formal statement of the faithfulness condition is as follows:

> **Faithfulness Condition:** [9, page 56,] Let $\mathcal{G}$ be a causal graph and $\mathcal{P}$ a probability distribution generated by $\mathcal{G}$. $< \mathcal{G}, \mathcal{P} >$ satisfies the

[2]Many of these properties of causes have been long known. Reichenbach described "causal forks" consisting of a cause and two or more effects. The effects are normally probabilistically dependent because of the common cause, but this dependence vanishes if we condition on the cause [6, page 158,]. The causal Markov condition is not completely uncontroversial. Salmon [6] postulates the existence of "interactive forks," that violate the causal Markov condition. Spirtes *et al.* give an appealing explanation of Salmon's examples and postulate that interactive forks do not exist, at least in the macroscopic world [9, Section 3.5.1,].

> Faithfulness Condition if and only if every conditional independence relation true in $\mathcal{P}$ is entailed by the Causal Markov Condition applied to $\mathcal{G}$.

One of the consequences of the causal Markov condition in combination with the faithfulness condition is conditional dependence: all causal predecessors of an observed variable v become probabilistically dependent conditional on v. Suppose that while riding a bicycle we get a flat tire. This makes all possible causes of the flat tire probabilistically dependent conditional on the flat tire. Observing pieces of glass on the road, for example, makes thorns less likely (the glass "explains away" the thorns).

Markov and faithfulness conditions bind causality with probability and along with other assumptions, such as acyclicity of the causal structure, reliability of the statistical tests applied, or independence of error terms, place constraints on the causal structure. The constraints provide clues to the causal structure that generated the observed patterns of interdependencies. Spirtes *et al.* show that given their assumptions, they are often able to reconstruct from a set of observations a unique causal structure of the system that generated them. The search for that causal structure is a search for the class of faithful models that are structurally able to generate the observed independences, and sometimes this search provides a unique structure.

4 Independence, Conditioning, Markov Condition, and Faithfulness

This section builds a bridge from the MBC to the ISC view of causality by introducing a deterministic notion of independence between a system's variables. This is a purely theoretical exercise that allows to talk about dependences among variables in a system of simultaneous structural equations. Please note that the concept of causal ordering, as explicated by Simon, operates on systems of simultaneous structural equations with no notion of uncertainty. Uncertainty enters these systems through variability of exogenous variables (error terms are simply exogeous variables on the par with other exogenous variables).

4.1 Deterministic Independence

I propose to base the deterministic definition of independence on the notion of dimensionality of the Cartesian product of variables. Followings the conventions in physics and mathematics, I define the dimension of a space roughly as the minimum number of coordinates needed to specify every point within it. A Cartesian product of n independent variables has dimensionality n, for, as each of the variables can vary independently over its domain, the points in this product cover an n dimensional space. If there is any interdependency among the variables, there will be loss in the dimensionality of this space. For example, if the element binding the two variables is an equation describing a unit circle, all we need to specify a point in this space is the polar coordinate angle. The Cartesian product of two independent variables forms a plane. If these two variables are dependent, then the domain of their Cartesian product will have a lower dimensionality and will be a line. The value of one of the variables puts a constraint on the value of the other.

Definition 1 (independence) *Sets of variables $\mathcal{X}$ and $\mathcal{Y}$ in a simultaneous equation model $\mathcal{S}$ are independent if the dimensionality of the Cartesian product of the variables in $\mathcal{X} \cup \mathcal{Y}$ is equal to the sum of dimensionalities of Cartesian products of variables in $\mathcal{X}$ and $\mathcal{Y}$ separately.*

Loss of dimensions is caused by functional relations that bind variables between the sets $\mathcal{X}$ and $\mathcal{Y}$. Each functional relation causes, in general, loss of one dimension. Because the exercise is theoretical, I will leave out of this paper the question how to test for deterministic independence in practice, along the lines of testing probabilistic independence.

4.2 Conditioning

Conditioning within a system of simultaneous equations means selecting a subset of observations that fulfills some specified condition. Such a condition forms a constraint on the values that a measured variable or a set of measured variables can take in the selected subset. Typically, one requires the value of a variable to be equal to some constant value. Conditioning is a passive way of "experimenting" with the system without modifying its causal structure. One selects those instances of the system's output that produce a specified value. If we condition on, for example, $x_i = x_{i_0}$, then we add to the system an additional constraint

$$x_i = x_{i_0} \ . \tag{1}$$

It is important to distinguish conditioning from direct manipulation of x_i, which is referred to in econometrics by *change in structure*. A change in structure is represented by replacing the equation that is made inactive by an equation describing the manipulation. In this case, one would replace the equation e_i that determines the value of x_i by the equation $x_i = x_{i_0}$. In conditioning, on the other hand, the selected data set needs to satisfy the equation e_i and, in addition, Equation 1. Conditioning on one variable reduces, thus, the system from a self-contained set of n equations with n variables to a set of $n+1$ equations with n variables (or, if we choose to replace x_i by a constant, n equations with $n-1$ variables), a system that is overconstrained. (This system still has solutions — these are the observed data points.)

4.3 Markov Condition

It turns out that in deterministic models, Markov condition can be derived rather than assumed and the following theorem can be proven.

Theorem 4 (Markov condition) *Let $\mathcal{S}$ be a simultaneous equation model with n variables $\mathbf{V}$ and n independent error variables. Let $\mathcal{G}$ be a directed acyclic graph with vertex set $\mathbf{V}$ reflecting causal ordering over variables $\mathbf{V}$ in $\mathcal{S}$. For every $w \in \mathbf{V}$, w is independent of $\mathbf{Z} \equiv \mathbf{V} \backslash (Descendants(w) \cup Parents(w))$ given $Parents(w)$.*

The theorem shows that the Markov condition is a simple consequence of the fact that the system is modeled by a set of simultaneous structural equations.

Theorem 2 shows that under the assumption of acyclicity, each of the equations determines one variable. Let equation e_i determine the variable x_i and precede (in the causal ordering over the model) all equations e_j, such that $i < j$. Because none of the equations e_k, such that $k < i$, contained x_i, each remains unchanged when we condition on x_i. Each of those equations e_k such that $i < k$ that contained x_i will now contain one fewer variable. This will lead to making the causal path from the predecessors of x_i to its causal successors inactive: note that as x_i becomes constant, none of the equations for the causal successors of x_i will depend on causal predecessors of x_i through x_i (they may, of course, depend through other paths).

4.4 The Faithfulness Assumption

I propose the following deterministic definition of faithfulness.

Definition 2 (faithfulness) *A structural equation model $\mathcal{S}$ is* faithful *with respect to its structure if and only if every independence between sets of variables in $\mathcal{S}$ is entailed by the structure of $\mathcal{S}$ (i.e., by the presence and the absence of variables in individual equations in $\mathcal{S}$).*

What this definition requires practically is that the model not contain equations that structurally look as if they were putting a constraint on a variable or a set of variables, but where in reality, the actual functional form and the actual values of the coefficients imply no constraint. Unfaithfulness may happen when a variable is present in an equation, but the coefficient of that variable is zero or becomes zero when influence through different paths is being computed (i.e., when the total effect of a variable on another variable through different paths "cancels out").

There are dependencies that do not result in loss of dimensionality, such as Peano or Sierpiński curves, or even the simple absolute value function. However, one has to remember that there are dependences that do not result in probabilistic dependence, for example deterministic dependences, excluded by the faithfulness axiom in the ISC approach.

4.5 Useful Properties of Causal Graphs

I report three properties of the relation between causal ordering and independence. Proofs are quite straightforward and omitted due to space constraints.

Theorem 5 (causal dependence) *If y precedes x in causal ordering, then y and x are dependent.*

Theorem 6 (spurious dependence) *If z precedes both x and y in the causal ordering, then x and y are dependent.*

One of most useful conclusions that can be drawn from conditioning is conditional dependence. Conditioning on a variable in a simultaneous equation model yields a data set in which all variables that are causal predecessors of that variable are dependent, contrary to the situation before conditioning, where exogenous variables in the system under study were independent by assumption. This observation shows that conditioning on a set of variables allows one to draw inferences about the causal ordering of variables, namely to discriminate, under certain circumstances, between causal predecessors and causal successors of the variables that were conditioned on. This property is captured by the following theorem.

Theorem 7 (conditional dependence) *Let $\mathcal{S}$ be a self-contained simultaneous equation model. Let Ψ be the set of causal predecessors of a variable x. Given the faithfulness assumption, any two subsets of variables $\mathcal{Y}, \mathcal{Z} \in \Psi$, are dependent conditional on x.*

The above three theorems show that causal ordering and interdependence are related. Causal ordering of the variables in a system of equations will result in a pattern of interdependencies in the observed data. This pattern, in turn, will give clues to the causal ordering or, more exactly, to the structural equations of the system.

5 Assumptions in Causal Discovery

Using elementary algebraic considerations, Simon [8] demonstrated that the problem of determining a causal structure, either from experimental or observational data, is severely underconstrained. The way one perceives the causal structure of a system is strongly dependent on the assumptions that one is willing to make. In particular, one might assume that a causes b only from an observed correlation between a and b, if one is willing to make the assumptions of time precedence and causal sufficiency (the latter excludes the possibility of a common cause) [8]. Similarly, one may be reluctant to accept even an experimental demonstration of causation, if one rejects critical assumptions about the experimental setup. It is, therefore, essential to state explicitly the assumptions made and provide the motivation for their validity.

In this section, I will outline the identifying information supplied by each of the assumptions made in causal discovery and the exact gains for discovering causality. I will use the structure matrix notation introduced in [2] and reproduced below to show the gains from each of the assumptions in terms of the number of coefficients of structural equations that are determined by the assumption. Some of the assumptions work in combination, and it is, therefore, difficult to assess the net gain obtained by each separately.

5.1 Initial Observations

The following definition, reproduced from [2], introduces a convenient notation for the structure of equations in simultaneous equation models.

Definition 3 (structure matrix) *The structure matrix A of a system $\mathcal{S}$ of n simultaneous structural equations $e_1, e_2, \ldots, e_n$ with n variables $x_1, x_2, \ldots, x_n$ is a square $n \times n$ matrix in which element a_{ij} (row i, column j) is non-zero if and only if variable x_j participates in equation e_i. Non-zero elements of A will be denoted by X and zero elements by 0.*

Our starting point is the observation that any system can be modeled by n measured variables ($x_1, x_2, \ldots, x_n$) and n unmeasured latent variables ($\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$), called error terms (note that I am not making any assumptions about their interdependence). If we denote the ith measured variable by x_i and the ith error term by $\mathcal{E}_i$, we can write the following structure matrix A for the set of $2n$ simultaneous structural equations with $2n$ variables. A solution of this set for any given set of values of the exogenous variables $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ describes a single observed data point.

$$\begin{bmatrix} & x_1 & \ldots & x_n & \mathcal{E}_1 & \ldots & \mathcal{E}_n \\ (e_1) & a_{11} & \ldots & a_{1n} & a_{1n+1} & \ldots & a_{12n} \\ (e_2) & a_{21} & \ldots & a_{2n} & a_{2n+1} & \ldots & a_{22n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ (e_n) & a_{n1} & \ldots & a_{nn} & a_{nn+1} & \ldots & a_{n2n} \\ (e_{n+1}) & a_{n+11} & \ldots & a_{n+1n} & a_{n+1n+1} & \ldots & a_{n+12n} \\ (e_{n+2}) & a_{n+21} & \ldots & a_{n+2n} & a_{n+2n+1} & \ldots & a_{n+22n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ (e_{2n}) & a_{2n1} & \ldots & a_{2nn} & a_{2nn+1} & \ldots & a_{2n2n} \end{bmatrix} \quad (2)$$

5.2 Acyclicity of the Causal Structure

The acyclicity assumption is probably the strongest assumption made in ISC causal discovery. It technically amounts to assuming that there are no feedback loops in the causal graph of the system. The implication of this assumption for a simultaneous structural equation model has been captured by Theorem 2. Every equation in such a model determines the value of exactly one endogenous variable.

Before showing the implications of the acyclicity assumption for causal discovery, I will rearrange the coefficients of the structure matrix to a form convenient in causal discovery. Without loss of generality we are free to assume that row i ($i = 1, \ldots, n$) in (2) represents equation e_i that determines the value of variable x_i, and row $n + j$ ($j = 1, \ldots, n$) represents equation e_{n+j} that determines the value of the error variable $\mathcal{E}_j$. Also, column i ($i = 1, \ldots, n$) of the matrix will contain coefficients for the variable x_i and column $n+j$ ($j = 1, \ldots, n$) will contain coefficients of the error variable $\mathcal{E}_j$. Mathematically this assumption amounts to rearranging the structure matrix by row and column exchanges (renaming the variables and the equations), which, as shown in [7], preserves the causal structure of the system.

The following three properties hold in this rearranged matrix A:

Property 1 (diagonal elements) $\forall_{1 \leq i \leq 2n}\, a_{ii} \neq 0$

Because I assumed that equation e_i determines variable x_i, and, therefore, x_i must be present in e_i, each diagonal element of A must be non-zero (i.e., $\forall_{1 \leq i \leq n}\, a_{ii} \neq 0$). The same holds for the error variables $\mathcal{E}_i$.

Property 2 (off-diagonal elements) *There are at least $2n(2n-1)/2$ zeros among off-diagonal elements of A. All non-zero off-diagonal elements in A represent direct causal predecessors of the diagonal element of the same row.*

In the proof of Theorem 2 [2], I demonstrate that another implication of the acyclicity assumption is that the structure matrix is triangular and, therefore, contains at least $2n(2n-1)/2$ zeros. The location of these zeros is only partly disclosed and can be retrieved only in combination with the properties of the observed data and other assumptions during the discovery process.

By Theorem 2, all variables that participate in an equation, except the one that is determined by the equation, are direct causal predecessors of that variable. By Property 1, the diagonal elements denote the variables that are being determined, therefore, it follows that all non-zero off-diagonal elements represent direct causal predecessors of the diagonal elements. Note that no assumptions have been made so far about interdependence of error variables and each of the equations e_{n+i} $(i = 1, \ldots, n)$ can model dependencies among these.

Property 3 (acyclicity) $\forall_{i \neq j}\, a_{ij} \neq 0 \Longrightarrow a_{ji} = 0$

$a_{ij} \neq 0$ implies that x_j is a direct predecessor of x_i and $a_{ji} \neq 0$ would imply that x_i is a direct predecessor of x_j, which then implies a cycle in the causal graph. Note that Property 3 captures only cycles of degree two. It is possible to capture cycles of higher degrees, although the conditions for these become increasingly complex.

5.3 Causal Sufficiency

The assumption of causal sufficiency[3] is equivalent to the assumption of independence of exogenous variables $\mathcal{E}_i$. Independence of exogenous variables amounts to assuming that half of the $2n$ equations contain just one variable, namely one of the n error terms. As the remaining n equations each involves exactly one

[3]This assumption can be relaxed in ISC causal discovery — some search algorithms proposed by Spirtes *et al.* allow for discovery of models that are not causally sufficient. In this case, the algorithm suggests possible common causal predecessors of any pair of the measured variables [9, Chapter 6,].

distinct variable of the n error variables, we get $3n^2 - 2n$ structural zeros.

$$\begin{bmatrix} & x_1 & x_2 & \dots & x_n & \mathcal{E}_1 & \mathcal{E}_2 & \dots & \mathcal{E}_n \\ (e_1) & X & 0 & \dots & 0 & X & 0 & \dots & 0 \\ (e_2) & a_{21} & X & \dots & 0 & 0 & X & \dots & 0 \\ (e_3) & a_{31} & a_{32} & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ (e_n) & a_{n1} & a_{n2} & \dots & X & 0 & 0 & \dots & X \\ (e_{n+1}) & 0 & 0 & \dots & 0 & X & 0 & \dots & 0 \\ (e_{n+2}) & 0 & 0 & \dots & 0 & 0 & X & \dots & 0 \\ (e_{n+3}) & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ (e_{2n}) & 0 & 0 & \dots & 0 & 0 & 0 & \dots & X \end{bmatrix}$$

For the sake of simplicity of the subsequent discussion, we can remove the error term parts of the above structure matrix (note that the removed parts contain no unknown values of parameters), obtaining:

$$\begin{bmatrix} & x_1 & x_2 & x_3 & \dots & x_n \\ (e_1) & X & 0 & 0 & \dots & 0 \\ (e_2) & a_{21} & X & 0 & \dots & 0 \\ (e_3) & a_{31} & a_{32} & X & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ (e_n) & a_{n1} & a_{n2} & a_{n3} & \dots & X \end{bmatrix} \tag{3}$$

In the structure matrix above, I have assumed that all zeros are located above the diagonal to show graphically the number of structural zeros obtained by the acyclicity assumption. In fact, the location of zeros is not disclosed a-priori and is only constrained by Property 3. The actual inference from the observed pattern of interdependencies concentrates on determining for each of the remaining $n(n-1)/2$ coefficients in (3) whether it is zero or non-zero.

6 Conclusion

Because the problem of causal inference from observations is severely underconstrained, the perceived causal structure depends on the *a priori* assumptions that one is willing to make. This paper has explicated the assumptions made in the causal discovery work (ISC view) and expressed them in terms of the earlier work in econometrics on structural equation models (mechanism-based view). I discussed the identifying strength of each of the assumptions in terms of the number of structural zeros and non-zeros that can be implied in the structure matrix.

The power of the ISC methods seems to rest on additional assumptions about causal relations that had not been made in econometrics. The two new powerful identifying assumptions are acyclicity of the causal structure and the assumption that each observed independence and dependence is a reflection of the causal structure of the system and is not merely coincidental (the latter called by Spirtes *et al.* "faithfulness assumption"). With respect to the faithfulness assumption, the new, previously unexplored, element is dependence of causes conditional on a common effect: a conditional dependence observed in this case is assumed to be structural and allows for distinguishing between predecessors and successors of the node conditioned on.

Acknowledgments

I am indebted to the late Herb Simon for his guidance and, especially, his contributions to my understanding of the mechanism-based view of causality. Had he been alive, he would have been a co-author on this paper. I thank Peter Spirtes, Richard Scheines, Clark Glymour, Chris Meek and Greg Cooper for inspiring discussions on the subject of causal discovery and helpful comments on this work.

References

[1] Gregory F. Cooper and Edward Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.

[2] Marek J. Druzdzel and Herbert A. Simon. Causality in Bayesian belief networks. In *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI–93)*, pages 3–11, San Francisco, CA, 1993. Morgan Kaufmann Publishers, Inc.

[3] William C. Hood and Tjalling C. Koopmans, editors. *Studies in Econometric Method. Cowles Commission for Research in Economics. Monograph No. 14.* John Wiley & Sons, Inc., New York, NY, 1953.

[4] Tjalling C. Koopmans, editor. *Statistical Inference in Dynamic Economic Models. Cowles Commission for Research in Economics. Monograph No. 10.* John Wiley & Sons, Inc., New York, NY, 1950.

[5] Judea Pearl and Thomas S. Verma. A theory of inferred causation. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *KR–91, Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452, Cambridge, MA, 1991. Morgan Kaufmann Publishers, Inc., San Mateo, CA.

[6] Wesley C. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, NJ, 1984.

[7] Herbert A. Simon. Causal ordering and identifiability. In Hood and Koopmans [3], chapter III, pages 49–74.

[8] Herbert A. Simon. Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, 49(267):467–479, September 1954.

[9] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Springer Verlag, New York, 1993.