Referee Report for "Learning Causal Relations in Multivariate Time Series Data."

This paper adapts results from the artificial intelligence literature to linear recursive structural equations systems with independently normally distributed error terms to provide necessary and sufficient conditions for observational equivalence among recursive "linear causal models." It then extends these results to the time series case. The paper also provides a simulation study and an empirical application.

My main comment on this paper is that it is not explicit in stating the assumptions underlying *d*-separation and the related observational equivalence results. This leads to inaccurate statements in some of the results of the paper. In particular:

- The paper first starts with employing results on "observationally equivalent" directed acyclic graphs (DAG). On page 4, the authors state that "the sparse DAG implies in particular a set of conditional dependence and independence among variables." However, the authors do not mention under which conditions are these implications correct. In fact, the factorization in equation (2.1) is not "implied" by figure 1. Instead this factorization is an assumption sufficient for *d*-separation to imply the conditional independence relations in statements (a) and (b) on page 4 for all compatible distributions with DAGs (a) and (b) in Fig.1. Pearl (2000, p.16) refers to this assumption as "Markov Compatibility." In addition, statement (c) on page 4 is not true for all distribution compatible with DAG (c) in Fig.1 but rather for at least one distribution compatible with DAG (c) (see Pearl 2000, theorem 1.2.4).

- On page 5, the authors state that "The rationale behind this assumption is that the basic features of a causal relation: transitivity, asymmetry, and non-reflexivity are well represented by a DAG." Absent a working definition of direct and indirect causality, it is not clear that these features are indeed "basic features of a causal relation." For example, transitivity of causal relations is not guaranteed if one adopts a definition of causal relations based on functional dependencies.

- In motivating the assumption of "stability" or "faithfulness," the authors state on page 7 that "the statistical procedure cannot differ whether this kind of independence is due to a particular chosen values of parameters of the underlying data-generating causal model or due to the causal independence of the underlying causal model." It's not clear what the authors mean by "the causal independence of the underlying causal model." I suggest rewriting this paragraph.

- Proposition 2.2 states that "a DAG model for X can be equivalently formulated as a linear recursive simultaneous equations model… ." A DAG does not assume linearity of the causal response functions. Indeed, the response functions can be nonparametric. Furthermore, one need not be concerned only with the conditional means when measuring causal relations. For example, it may be that the variance or a certain quantile of the response variable is causally affected by the cause of interest where as the mean of the response variable is not. The authors should clarify whether linearity holds under

normality of the error terms when interest attaches to other features of the distributions of the response variable other than the mean.
In addition, proposition 2.2 claims the equivalence of a DAG and a linear recursive simultaneous equations model where as Remark 1 on page 8 states that "from a DAG of jointly normally distributed variables we may sometimes get different linear recursive simultaneous equations models." The authors should reconcile these two claims. Also, I suggest the authors define Σ in proposition 2.2 and explain what they mean by a "symmetric DAG" on page 8.

- In Remark 2 on page 9, it is not clear to me what does "the DAG with the most explicit conditional independence" mean.

- Definition 2.3 states that "if two linear causal models can always generate identical joint distribution, they are called observationally equivalent." Under certain assumptions, two DAGs are observationally equivalent if they encode the same set of conditional independence relationships. They need not "generate an identical joint distribution."

- Remark 3 on page 11 states that "the change in the direction of the arrow $x_i \rightarrow x_j$ will not lead to a cycle." In fact, I am not aware of any results that preclude a cyclic and acyclic "causal models" of being observationally equivalent. Rather the assumption of acyclicality is imposed to limit the search to the class of observationally equivalent acyclic models.
Also, in Remark 3 on page 11, the authors state that "we can alter the direction of the arrow $x_i \rightarrow x_j$ to get an observationally equivalent model if x_i and x_j have the same parents. The sentence needs to be rephrased since the presence of the arrow $x_i \rightarrow x_j$ in a DAG implies that x_i is a parent of x_j and hence x_i and x_j can not have the same parents.
In addition, on page 11 following Remark 3, the authors state that "other direction of edges in DAGs do not have any causal implication." The sentence should clarify what is meant by "causal implication." Does it mean implications on the resulting set of conditional independence relationships and hence on the class of observationally equivalent acyclic causal models?

- On page 17, the authors state that "zero elements in the coefficient matrices A_i implies corresponding causal independence." I suggest the authors define the term "causal independence."

- A key feature underlying the "Causal Markov" assumption and driving the results concerning *d*-separation as well as the results of section 2 and 3 in this paper is that the error terms in the structural equations are independent and identically distributed. This rather strong assumption is also crucial to the results concerning Granger Causality in section 4. Among other things, it implies the absence of unobserved latent variables in the model. In fact, the IC algorithm is no longer valid when latent variables are present. It is helpful if the authors discuss more explicitly the role that the assumption of i.i.d error terms plays.

- The authors do not discuss the results in the fifth column of Table 1.

- In the empirical application, the greedy search algorithm yields a specific DAG containing information on direct and indirect causal relations among the variables of interest as depicted in Figure 3. The authors do not comment on the economic content of this DAG. Are these causal relationships plausible? How do they relate to the literature on wage-price dynamics? What are the economic implications of the assumptions of absence of latent variables, independence of the error terms, and homoskedasticity?
Also, the authors do not explain why they restrict their sample to the range 1965:1 to 2004:4.

- Miscellaneous comments:
    - I find the paragraph following proposition 2.1 hard to read.
    - Page 9, third paragraph, first sentence: the order of listing conditional covariance and conditional variance does not match the order of listing their corresponding symbols.
    - Typo: page 13, first sentence in section 3: should be "inferring causal relations."
    - Typo: proposition 4.2, second bullet, last sentence: should be "causal" and not "casual."
    - Typo: page 36, proof of lemma 7.1: need to add a bracket after $Z = \{z_1, \ldots, z_m\}$ and the parenthesis after $P(x \mid z)$).
    - Typo: page 38, third to last sentence: should be $a^*_{k+2,k+1}$.

In Conclusion, I view the attempt to adapt results on causal model selection from the artificial intelligence literature to the familiar structural equations setup to be worthy. I think this paper would be most helpful if it explicitly states the assumptions imposed and focuses on the time series extension and primarily on the simulation and empirical application.