1 Responses to the Referee Report III

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. -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).

Response:

d-separation is an important graph criterion to judge the conditional independence among variables. It plays a very important role in the learning algorithms, such as PC, IC or IC*, to uncover the causal relations from data. However important the d-separation may be, it is not necessary to use d-separation to present the theory of inferred causation. Even in the fundamental paper of Pearl and Verma (1991): "A Theory of Inferred Causation", the word d-separation is not mentioned.

The fundamental assumption of the method of inferred causation is that, as given in Definition 2 in Pearl and Verma (1991): the casual relations among a set of variables U can be modelled in a dag D and a set of parameters Θ_D, compatible with D. Θ_D assigns a function $x_i = f(pa(x_i), \epsilon_i)$ and a probability measure g_i to each $x_i \in U$, where $pa(x_i)$ are parents of x_i in D and each x_i is a random disturbance distributed according to g_i independently of the other ϵ's and of any preceding x_j: $0 < j < i$.

The probability measure compatible with D is called to satisfy the Markov condition in Pearl (2000). The Markov condition implies in particular that the disturbance ϵ_i are independent form other ϵ's. In addition to the Markov condition, the minimality of the causal structure, D, and the stability of the distribution are two key assumptions on the data-generating causal model to rule out the ambiguity of the statistical inference in recovering the data-generating causal models.

Following the comments of the referee, we state explicitly the three assumptions in section 2.1 and adopt the definition of causal model from Verma and

Pearl (1991). We formulate the statements (a) (b) and (c) on page 4 in terms of d-separation.

- 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.

Response:

Transitivity, asymmetry and non-reflexivity are three features mentioned in Spirtes, Glymour, and Scheines (2001) to motivate an axiomatic definition of causal models. As correctly pointed out by the referee, without a working definition of causality it is not clear that these features are indeed "basic features" of causal relation. Hence, we leave out this paragraph.

- 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. Its not clear what the authors mean by the causal independence of the underlying causal model. I suggest rewriting this paragraph.

Response: We rewrite this paragraph.

Given a set of data generated from a causal model, a statistical procedure can principally identify all the conditional independence. However, the statistical procedure cannot differ whether this kind of independence is due to a lack of the edge in the DAG of the causal model or due to particularly chosen parameter values of the DAS such that the edge in this case implies the independence. To rule out this ambiguity, Pearl (2000) assumes that all the identified conditional independence are due to lack of edges ind the DAG of the causal model. This assumption is called stability condition in Pearl (2000). In Spirtes et al. (2001) it is called faithfulness condition. This assumption is therefore important for interpreting the conditional dependence and independence as causal relations.

- 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 in proposition 2.2 and explain what they mean by a symmetric DAG on page 8.

Response:

As correctly pointed out by the referee, a DAG does not assume linearity of the causal response functions. In our paper we consider only jointly normally distributed variables X, as in most application in macroeconomics. In this case we have the the linearity: the causal response functions is linear and the disturbances are independent and normal. Our results hold under this condition but not for arbitrary distributed X.

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 in proposition 2.2 and explain what they mean by a symmetric DAG on page 8.

Response: We reformulate the paragraph with an example.

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

Response:

This should mean "the minimal structure". We reformulated the sentence.

- *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.*

There are two kinds of model equivalence: dependence equivalence, as stated by the referee, and the distributional equivalence. In DAG models with jointly normally distributed variables X, these two equivalences are the same.

- *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. Ih In fact, I am not aware of any* *results that preclude a cyclic and acyclic gcausal modelsh 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 xi 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?*

Response:

Following the comments of the referee, we reformulated the paragraph. Remark 3 presents an example for Proposition 2.1.

- 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.*

Response:

Following the suggestion of the referee we define the causal independence between x_i and x_j as no edge between them.

- 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.

Response:

As correctly pointed out by the referee the key assumption is the Markov condition. It implies that the error terms are independent from each other. But they do not need to be identically distributed.

Surly the existence of latent variables is an important issue in the theory of inferred causation. IC algorithm can be even used to detect the existence of latent variables. However, the input of the IC algorithm is a valid estimate of the covariance matrix. Even in case of missing variables, the estimates of pairwise covariances of the observables are still valid. Therefore the input estimate of the covariance matrix of the observables is valid.

But in the case of time series models, the existence of missing variables will leads to invalide estimate of the residuals matrix, which will be used as input of learning algorithm. An invalid estimate of the covariance matrix will not lead to an useful results. Therefore we have to assume the casual sufficiency in our paper.

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

Response: we add a comment on 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.

We add comments on the recovered DAG, and discussion the issues mentioned above.

- Miscellaneous comments:

- I find the paragraph following proposition 2.1 hard to read.

Response:

We reformulate the paragraph.

- 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.

Response: A correction is made.

- 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 = z1, , zm and the parenthesis after P(x — z)). - Typo: page 38, third to last sentence: should be $a*_{k+2,k+1}$.

Response: We made the corrections.

2 Responses to Referee Report II

Major Remarks

My only concern about this paper is on the assesment of the performance of the methodology developed in the paper. The paper considers a simulation exercise for this purpose as in Demiralp and Hoover (2003). That is they generate data from a variety of known specifications of SVARs and then address the question of how successfully A0 can be recovered from estimates of VARs. However, the problem for empirical analysts is to evaluate the reliability of such identifications when A0, and, indeed, the entire specification of the SVAR is unknown. To address this question, Demiralp, Hoover, and Perez (2007)1 employ a bootstrap strategy. Starting with the original data, they estimate the VAR and retain the reduced form residuals, t = 1, 2, . . . , T.. In order to maintain the contemporaneous correlations among the variables, they resample the residuals by columns from . The resampled residuals are used in conjunction with the coefficient estimates of the VAR to generate simulated data. A large number of simulated data sets are created. For each one, they run the search algorithm, record the results, and compute summary statistics. A similar exercise can be considered to evaluate the performance of the technique developed in this paper, perhaps in a follow up paper.

Response:

We agree totaly with the referee in that alternative methods should be used to explore the performance of the 2-step procedure to learn causal relations proposed in our paper. The bootstrap method used in Demiralp, Hoover and Perez (2007) is surely a useful one to assess the performance of the procedure. We plan a follow up paper to carry out this exercise, as suggested by the referee.

Minor Remarks 1) On page 8, a directed acyclic graph (DAG) is told to be equivalent to a simultaneous equation model (SEM). However, an SEM allows for a circular feedback whereas a DAG does not. Wouldnt a DAG rather correspond to a a seemingly unrelated regression (SUR) model? Response:

As correctly pointed out by the referee, SEM may allow a circular feedback, therefore DAG corresponds to a **recursive** SEM with independent errors. A SUR with correlated errors, can be transformed into a recursive SEM with uncorrelated errors through, for example, orthogonalization of the errors.

2) On page 14 (under Remarks), it is told that if the significance of the test converges to zero, as the number of observations goes to infinite (emphasis added). I believe zero should be replaced by one.

Response: The significance level of a test is the probability to reject the null when then null is true. Therefore that the significance level goes to zero means the probability to identify the true model goes to one.

3) At the end of p. 16, there is an expression $i > p$, and I could not find the definition for p.

Response: p is an assumed lag-length.

4) In footnote 21, it is admitted that the choice of one lag using SIC is very unusual. Did the authors consider an alternative lag selection criteria such as AIC, and are the results robust to lag lenght? My experience is that AIC and SIC do not necessarily agree on the lag lenght and I am curious to know if the results show any sensitivity

The choice of leg length depends on the criteria used. AIC suggests usually a longer lag than SIC. In our example it happens that AIC, BIS and also the likelihood ratio test suggest a lag length of one.

3 Responses to the Referee Report I

The paper presents a rigorous approach to identifying the causal struc- ture underlying multivariate time series data and linking this approach to the structural VAR methodology. Since structural VARs have pro- liferated in the Macroeconomics literature as a way of examining the implications of alternative economic models or hypotheses, the paper provides a useful contribution in terms of bridging the gap between the method of inferred causation based on graph theoretic notions and VAR methodology.

We agree total with the comments of the referee.

The paper does not appear to provide new results on the probabilistic causal approach. Instead it relies on Pearl (2000) and others for this purpose.

Response:

Also, here we agree with the comments of the referee. One important issue is this respect ist that the causal model presented in this paper cannot take simultaneity into account, though the notion of simultaneity is very important in economics. Therefore, much more work has to be done to develop a causal model that may handle the simultaneity issue.

The papers contribution is to make the link from the method of inferred causation and the DAG to simultaneous equation models (SEMs) and thence to time series models. This discussion is quite clear and allows the reader to understand the relation between these approaches. One of the issues that I found lacking in the current approach is a closer link to economic theorizing. Even in the economic application regarding the wage-price spiral, the analysis was presented in terms of relatively atheoretic Phillips curves. Given the tremendous advances made in the macroeconomics literature in terms of modelling macroeconomic phenomena, the argument that the Phillips curves had been derived based on data-driven causal analysis was less than satisfactory for me.

Response:

This is surely a very challenging comment, which has also been raised in the Referee Report III. We view the main purpose of this paper as providing a methodology to uncover the regularity in data. The obtained contemporaneous and temporal causal structure allows a theoretical interpretation. We have added this in the paper. However, our main interest here is to see whether the data-driven results are compatible with the results obtained by theoretic reasoning and whether this method can lead to new insight of the problem.

Another question that came to mind was the relation of this approach to dynamic factor analysis. Much recent work in empirical Macroeco- nomics has been concerned with identifying a small number of shocks underlying cyclical phenomena. Giannone, Reichlin and Sala (2006) show how more general classes of equilibrium business cycle models can be cast in terms of the dynamic factor representation. They also describe how to derive impulse response function for time series mod- els which have reduced rank, that is, ones for which the number of exogenous shocks is less than the number of series. It would have been of interest to see how the current approach relates to dynamic factor models more generally.

Response:

The main purpose of dynamic factor models is to reduce the dimensionality of the data und to condense the information. The causal analysis does not intent to condense the information but to uncover the possible casual structure in the data. Further, a causal model provides an ordering of the variables in the model according to causal direction. The focus of the analysis is on the variables themselves. The dynamic factor model aims at detecting the underlying factors that drive the variables. The focus of the analysis is the connection between the factors and the variables. Sofar there is no obvious link between these two models. However, a more detailed analysis may shed more light on this issue.

The paper certainly has enough material but does it have too much? By the time the reader gets to Section 5, s/he is loaded down with alter- native models and concepts. Would a re-organization of the paper help the reader, especially the more empirically oriented one? For example, one approach would be to present the application first and note that standard Granger causality analysis leaves an ambiguity regarding the causality structure underlying wage-price dynamics. This substantive issue could then be used to motivate the relationship among DAGs, SEMs and more specifically, structural VARs.

Response:

We understand our paper mainly as a methodological paper. The analysis with empirical data should only provide a illustrative example for application of this method. Also, the relation of the causal model to the Granger causality is not the main focus of the paper. Therefore, we present at first the concept of inferred causation and then the extension to time series data and as last an application.

References

PEARL, J. (2000). *Causality*. Cambridge University Press, 1st edition.

SPIRTES, P., GLYMOUR, C., AND SCHEINES, R. (2001). *Causation, Prediction and Search*. Springer-Verlag, New York / Berlin / London / Heidelberg / Paris, 2nd edition.