Referee Report for "Learning Causal Relations in Multivariate Time Series Data"
by Pu Chen and Hsiao Chihying

This paper presents a nice application of graph theory techniques for the identification of a dynamic causal model (TSCM). Starting with Swanson and Granger (1997), a number of authors have demonstrated how graph-theoretic techniques, can in some cases use the data themselves to identify the SVAR. These methods exploit patterns of conditional independence in the data. In cases in which unique identification is not possible, they may nonetheless reduce the class of admissible identifying assumptions considerably. Graph-theoretic search algorithms work according to the following general plan: the true contemporaneous correlations matrix $\mathbf{A}_0$ in a VAR induces a set of conditional independence relations among the elements of reduced form errors. The algorithm thoroughly tests for conditional independence relations among the estimated reduced form errors. It then selects the class of Choleski factors – perhaps unique – that is consistent with those independence relations.

The main contibution of this paper is to extend the previous econometric applications of graph-theory in a dynamic set-up. That is, they identify a structural VAR in terms of a contemporaneous causal structure *as well as* a temporal causal structure. Adding the temporal causal structure is an important contribution to this recently developing literature.

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 $\mathbf{A}_0$ can be recovered from estimates of VARs. However, the problem for empirical analysts is to evaluate the reliability of such identifications when $\mathbf{A}_{0,}$ 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 $\hat{\mathbf{U}} = [\hat{\mathbf{U}}_t]$, t = 1, 2, . . . , T.. In order to maintain the contemporaneous correlations among the variables, they resample the residuals *by columns* from $\hat{\mathbf{U}}$. 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.

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. Wouldn't a DAG rather correspond to a a seemingly unrelated regression (SUR) model?
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.
3) At the end of p. 16, there is an expressin i>p, and I could not find the definition for p.
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.

[1] Demiralp, S., Hoover, K. and Perez, S. (2007) "A Bootstrap Method for Identifying and Evaluating a Structural Vector Autoregression," unpublished manuscript