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

This paper uses graph theory to discuss the issue of eliciting causal relations in multivariate time series data. The application of the method of inferred causation to learn the causal relation in the observed data is the topic of study by Pearl (200), Heckerman, Geiger, and Chickering (1995) and others. The current paper relates this approach to develop causal models for multivariate time series data. The paper is comprised of five parts. Part 1 is a brief introduction. Part 2 lays out the notion of inferred causation while Part 3 briefly discusses methods for uncovering causal relations using alternative algorithms. Section 4 relates these concepts to time series models and Section 5 presents an application to wage-price dynamics.

The main idea behind the method of inferred causation is that causal relations among a set of variables can be represented by a sparse directed acyclic graph (DAG). In such a graph, the vertices represent the variables and a directed edge connects a causal variable to an effect variable. A *causal model* is then a DAG together with the conditional distributions based on the orderings implied by the DAG. The *learning* of the DAG is the method of uncovering the data generating the DAG. Consider a DAG consisting of n variables, $(x_1, \ldots, x_n)'$. A consistent model selection criterion evaluates the model in terms of the sum of its likelihood and a penalty for the dimensionality of the model. Thus, statistically learning the causal order among the variables implies searching for the most parsimonious model that can account for the joint distribution of $(x_1, \ldots, x_n)'$. Issues that must be addressed when following this approach include the observational equivalence of alternative

1

DAG models, and whether the conditional independence in a given set of data is due to particular chosen values of the parameters or to the causal independence underlying the causal model. The authors rely on the results of Pearl (2000) for the definition of observationally equivalent classes of DAG models, and also for assuming that all the identified conditional independence is due to the causal independence of the underlying causal model. The remainder of Section 2 in the paper is then devoted to relating these notions to linear recursive simultaneous equation models (SEM's). In particular if a vector $(x_1, \ldots, x_n)'$ is jointly normally distributed, it immediately follows that a DAG is equivalent to a SEM. (See Proposition 2.2.) An SEM is characterized by a lower triangular coefficient matrix with 1's on its diagonal. A series of propositions delineates the conditions under which the causal order among the variables can be inferred from the SEM. Section 3 discusses heuristic algorithms for inferring the causal model from a given set of data.

The main results of the paper are in Section 4. This section extends the earlier notions to multivariate time series data. Consider a sequence of n-vectors $\{X_t\}$ for $t \in I$ where I is some index set. In graph theory parlance each vertex of a DAG now corresponds to a random element X_{it} for $i = 1, \ldots, n$ and $t \in I$. Since there is only one observation on X_{it} in a typical time series application, many restrictions have to be imposed on the recursive model to allow statistical inference. One restriction is a temporal causality or measurability constraint, namely, that X_t cannot be a cause for $X_{t-\tau}$ for $\tau > 0$. A second plausible restriction is the invariance of the causal relation between X_{t+k} and $X_{t-\tau+k}$ for any $k > 0$. Finally, a third restriction is that the causal influence from past X's to current X depends only on a finite number of lags p. A linear recursive model of time series that satisfies these three restrictions is called a *time series causal model (TSCM)*. (See

Definition 4.1.) Under these assumptions, a TSCM can be expressed as:

$$A_0 X_t + A_1 X_{t-1} + A_2 X_{t-2} = \epsilon_t, \quad t = p+1, \ldots, T,$$

where $E(\epsilon_t \epsilon'_{t-s}) = 0$ and $E(\epsilon_t \epsilon'_t) = D$ where D is a diagonal matrix. The causal relations among the variables are expressed in terms of the restrictions on the coefficient matrices. A_0 is lower triangular and it describes the contemporaneous causal relations among the elements of X_t. The matrices A_i describe the causal dependence between X_t and X_{t-i}, with zero elements in the coefficient matrices A_i corresponding to causal independence. In Proposition 4.2, the relationship between Granger causality and a TSCM is established. What is required for Granger non-causality of $X_{k,t}$ on $X_{i,t}$ is that $X_{k,t-s}, s = 1, \ldots, p$ does not have a temporal causal influence on $X_{i,t}$ and also on the predecessors of $X_{i,t}$ in the TSCM. However, the paper shows that the absence of probabilistic causation does not imply Granger non-causality. The remainder of Section 4 is devoted to methods for recovering the TSCM from the data. A two-step procedure is proposed whereby the contemporaneous causal structure summarized by A_0 is uncovered using the algorithms such as the Greedy Search algorithm and the temporal causal structure summarized by A_i is uncovered using the BIC criterion. Section 4.4 conducts simulation studies to determine which causal structures can be uncovered using the two-step procedure outlined earlier. Finally Section 5 presents an application of this approach to the analysis of wage-price dynamics.

Comments:

- The paper presents a rigorous approach to identifying the causal structure underlying multivariate time series data and linking this approach to the structural VAR methodology. Since structural VAR's have proliferated in the Macroeconomics literature as a way of examining the

3

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.

- 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.

- The paper's contribution is to make the link from the method of inferred causation and the DAG to simultaneous equation models (SEM's) 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.

- Another question that came to mind was the relation of this approach to *dynamic factor analysis.* Much recent work in empirical Macroeconomics 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 models which have reduced rank, that is, ones for which the number of

4

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.

- 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 alternative 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 DAG's, SEM's and more specifically, structural VAR's.

References

Giannone, D., L. Reichlin and L. Sala (2006). "VAR's, Common Factors and the Empirical Validation of Equilibrium Business Cycle Models, *Journal of Econometrics* 132, 257–279.