REMARKS CONCERNING GRAPHICAL MODELS FOR TIME SERIES AND POINT PROCESSES

David R. Brillinger [*]

Resumo

Uma rede estatĭstica é uma coleçăo de nós representando variáveis aleatórias e um conjunto de arestas que ligam os nós. Um modelo estocástico por isso e chamado um modelo gráfico. Estes modelos, de gráficos e redes, sáo particularmente úteis para examinar as dependéncias estatĭsticas baseadas em condiçoes do tipo das que ocorrem frequentemente em economia e estatĭstica. Neste artigo as variáveis aleatórias dos nós serão séries temporais ou processos pontuais. Os casos de gráfos direcionados e năo-direcionados são apresentados.

Abstract

A statistical network is a collection of nodes representing random variables and a set of edges that connect the nodes. A probabilistic model for such is called a graphical model. These models, graphs and networks are particularly useful for examining statistical dependencies based on conditioning as often occurs in economics and statistics. In this paper the nodal random variables will be time series or point proceses. The cases of undirected and directed graphs are focussed on.

1. Introduction.

Graphical models, that is probability models on networks of vertices and edges, are experiencing a surge of research development today. One can point to the books by Whittaker (1990) and by Edwards (1995) as well as to many papers, for example those listed in the References section. The study of graphical models has become important in the field of statistics and in the social sciences generally. It can be anticipated that such models will prove useful in addressing basic questions arising in economics concerning structural and causal modelling.

Given its historical roots in path analysis, simultaneous equations, structural and finally recursive models, research on graphical models is of importance to economists and econometricians. An economy can be viewed as a vast system or network of interconnected processes. In various circumstances the connections (or edges) are of particular interest and the processes are either time series or point processes. Structural approaches, eg. via time series or econometric models, are well known to economists and econometricians, see eg. Granger (1980), Hendry and Richard (1983) and Harvey (1989). This paper focuses on some of what graphical models have to offer.

As an example of a conceptual economic network, of some historical interest, one can mention the Philips machine, Philips (1950), Swade (1995). Pictures of the machine may be found in those references. Figure 1 presents a graph of Philips' Simple Model. The machine was meant to represent the behaviour of the British economy in the late forties. The arrows indicate the direction of the flow of liquid in the machine. Water is pumped to the top and flows downwards. Savings (S) are siphoned off into a separate loop and fall into a tank (M'). Portions of these rejoin the main flow as expenditure (E) and investment (I). The net flow at the bottom accumulates in a tank. The level of the tank represents the working balance (M) required for a given level of economic activity. The liquid is pumped back to the top. The flow of water

represents the movement of money and the accumulation of money is represented by the water collecting in tanks. Results are read off from scales located at various places. Flow is initiated, for example, by altering expenditure. The series involved here are (continuous) time series.

Less often economic data of point process form, $\{\tau_j\}$, occur. That is the measurements are the times of occurence of certain events, such as the times of change of a country's prime interest rate or the ticker times of stock sales. Economic data may also be of marked point process form, consisting of times and associated values (marks), eg. the times of change of the prime rate together with the value of the new rate, $\{(\tau_j, M_j)\}$. In this paper time series and point processes are labelled by the nodes of a graph and corresponding statistical models discussed. The models may involve the processes at vertices u and v being conditionally independent given the remaining vertices. Such a question may be formulated as involving the existence or nonexistence of an edge connecting vertices u and v. Recognizing such structure allows decomposition of the estimation problem into simpler components, in various cases. The question of measuring the strengths of connections present is also addressed in this article.

The use of graphs to represent statistical models dates back at least to Wright (1921) and his work with path models. One can also mention the graphs of Tinbergen (1939), page 138, for causal connections between disturbances and profits and of Wold (1956), page 45, for a demand-supply model. Another example is provided by the schematic diagrams so often employed in systems analysis. Networks are present in compartment analyses (see Jacquez (1972)). Here substances flow between and are retained in certain compartments. Influence diagrams and Bayseian networks are related topics, see Smith (1989), Normand and Tritchler (1992). Pearl (1995) presents a framework for causality and the manipulation of probabilities associated with such graphical models. Causal analysis is further discussed in Glymour et al. (1987). An example of a marked point process is provided by a queuing network. Here there is a

set of interconnected nodes, each node consisiting of a queue, where customers wait for service, and there are one or more servers. (See Brémaud (1981), for example.)

This present work consititutes a simple beginning discussion of the time series and point process cases of graphical models. It is to be noted that the meanings of graphical representations differ between authors and applications, and this needs to be kept in mind when reading the literature.

The layout of the paper is: first some basic concepts of graph theory and graphical models are presented, then consideration turns to the cases of continuous and discrete ordinary random variables, finally the time series and point process cases are addressed. The paper ends with a discussion and summary.

2. Graphical Models.

A *graphical model* is a statistical model embodying a set of conditional independence relationships which may be summarized by means of a graph. This section presents some general concepts and then moves on to the specific cases of undirected and directed graphs. The goal, for inferential purposes, is the likelihood function. General references include: Whittaker (1990), Lauritzen et al. (1990), Wermuth and Lauritzen (1990), Cox and Wermuth (1993), Edwards (1995).

A *graph* is a pair $G = (V,E)$ where V is a finite set of nodes (or vertices) and E, the set of edges, is a subset of $V \times V$, the ordered pairs of distinct nodes. Figure 2 provides an example with 4 nodes and 2 edges. An edge (α,β) in E is called *undirected* if both (α,β) and (β,α) are in E. An undirected edge is indicated by a line. A graph is *simple* if there are no multiple edges or loops. (This will be assumed throughout the paper.) This is the case in Figure 2. A *clique* is a maximal set of nodes, joined to each other. For Figure 2 the cliques are: {1}, {2,4}, {3,4}. An *undirected graph* is one with only undirected edges. An edge (α,β) is called *directed* if (α,β) is in E, but not (β,α). A directed edge is denoted by an arrow $\rightarrow$. A

directed graph is one in which all edges are directed. Figure 3 and 4 provides examples of directed graphs. There are two directed edges in Figure 3 and three in Figure 4. An *acyclic directed graph* has no paths beginning and ending at the same node. The graphs of Figure 3 and 4 are acyclic, while the graph of Figure 1 is cyclic (feedback is present).

2.1 Undirected Graphs.

A set B of vertices *separates* the sets A and C if all paths from vertices in A to C pass through B. An undirected graph possesses the *global Markov property* if for any triple (A, B, C) of disjoint subsets of V, such that B separates A from C, the variate $(Y_a)_{a \ in \ A}$ is statistically independent of the variate $(Y_c)_{c \ in \ C}$ given the variate $(Y_b)_{b \ in \ B}$. Conditions for this occuring may be found in Lauritzen et al. (1990). In broad circumstances it may be shown that the probability density of $\mathbf{Y}$ then has the form

$$f(\mathbf{y}) = \prod_{\gamma} \psi_\gamma(\mathbf{y}) \tag{2.1}$$

where the product is over all the cliques γ of G. For the graph of Figure 2 this means

$$f(y_1, y_2, y_3, y_4) = g(y_1)h(y_2, y_4)i(y_3, y_4)$$

Here the variates Y_2 and Y_3 are independent given the variate Y_4. This type of conditional independence is typical of undirected graphs. Taking $f(.)$ to be say a multivariate normal density, the factorization can indicate directly parameter values which need to be set to 0.

For the case of undirected graphs, Geiger and Pearl (1993) show that every axiom for conditional independence is an axiom for graph separation and that every graph represents a consistent set of independence and dependence constraints. They conclude that graphs provide a safe language for encoding statistical dependencies.

There is also a form of the results for discrete variates, with the densities replaced

by probability mass functions.

2.2 Directed Graphs.

Consideration next focuses on directed graphs in the acyclic case. In empirical analyses the arrow may represent a direction of influence or of causation or of a variate value at an earlier time point. For a and b vertices, a is a *parent* of b if there is a directed edge $a \rightarrow b$. The set of parents of the vertex b is denoted $pa(b)$.

A directed acyclic graph possesses the *directed global Markov property* if the density admits a recursive factorization,

$$f(\mathbf{y}) = \prod_{v \ in \ V} k_v(\mathbf{y}_v, \mathbf{y}_{pa(v)}) \qquad (2.2)$$

for some functions k_v. The term $k_v(\mathbf{y}_v, \mathbf{y}_{pa(v)})$ has the interpretation of the conditional density of $\mathbf{Y}_v$ given $\mathbf{Y}_{pa(v)}$. Conditions for this property are given in Lauritzen et al. (1990).

For the graph of Figure 3 the factorization is:

$$f_1(y_1)f_2(y_2)f_3(y_3)f_4(y_4|y_2,y_3)$$

with $pa(4) = \{2,3\}$. For Figure 4 it is

$$f_1(y_1)f_2(y_2|y_1,y_3)f_3(y_3|y_1)$$

with $pa(2)=\{1,3\}$ and $pa(3) = \{1\}$. One might consider the hypothesis

$$f_2(y_2|y_1,y_3) = f_2(y_2|y_1)$$

i.e. that there no direction connection of vertex 3 to vertex 2. The expressions above lead directly to likelihood functions useful for inferential purposes.

2.3 Strength of connections.

Suppose there is a graphical model. One may wish to describe the strength of connection associated with a particular edge. Two (related) measures may be considered. The first is the *mutual information*, the second is the *change in deviance* when an edge is removed from the model. The measures will be seen to be quite

related.

Consider continuous random variables X and Y with joint density $p(x,y)$ and marginals $p_X(x)$ and $p_Y(y)$. The mutual information between X and Y is

$$I_{XY} = \iint p(x,y) \, \log \frac{p(x,y)}{p_X(x)p_Y(y)} dxdy$$

It is seen to be 0 in the case that X and Y are independent and > 0 otherwise. For some discussion of this quantity see Granger and Hatanaka (1964), Kanter (1979) and Parzen (1983).

The mutual information may be related to the differential entropy of a random variable. For a random variable Y, with density $p_Y(y)$, the differential entropy is

$$H_Y = -\int p_Y(y) \, \log p_Y(y) \, dy$$

and one sees that

$$I_{XY} = H_Y - H_{Y|X}$$

in particular.

The *deviance* of a stochastic model is minus twice the loglikelihood, $-2 \log p_Y(y)$. The change in deviance that results from fitting a submodel proves a useful quantity. Consider fitting the submodel of independence, $p(x,y) = p_X(x)p_Y(y)$, then the change in expected deviance is

$$-2\iint p_X(x)p_Y(y) \, \log p_X(x)p_Y(y) \, dxdy + 2\iint p(x,y) \, \log p(x,y) \, dxdy$$
$$= 2 \, I_{XY}$$

Deviance change and mutual information are thus closely related and either may be considered as a measure of strength of a connection.

In the case of a graphical model one can study the change in deviance when one fits, both including a particular edge and not. The values may be added to a graph in a presentation of the results of the analysis. The edge might be directed or not. Some computer programs are available for carrying out the computations, eg. MIM of

Edwards (1995).

3. The Continuous Case.

Suppose one is dealing with random variables having continuous distributions.

3.1 The Undirected Case.

In the Gaussian case, conditional independence corresponds to zeroes in the inverse of the covariance matrix. Such models were studied in Dempster (1972) and are referred to as covariance selection. The model on the graph will have the conditional independences built in and estimates will need to reflect this. The inverse matrix corresponds to an undirected graph in the sense that the entry in row u, column v being 0 corresponds to no edge (u,v) in the graph whose nodes are all the labels. More specifically let $\{Y_v, v \, \varepsilon \, V\}$ be normal with covariance matrix Σ. Suppose vertices u and v are unconnected in the independence graph. The variates Y_u and Y_v are then statistically independent given the rest and

$$(\Sigma^{-1})_{uv} = 0 = R_{uv \mid rest}$$

where $\mathbf{R}$ is the matrix of partial correlations. See Whittaker (1990) for a development of this result.

To examine the possibility of 0 partial correlation empirically, one can look for near 0 entries in an estimate $\hat{\mathbf{R}}$ of $\mathbf{R}$ or examine the change in deviance resulting from assuming first $R_{uv \mid rest}$ is arbitrary and then that it is 0.

3.2 The Directed Case.

One would work with joint densities factorizing as in (2.2) and examine the difference in deviance between the models with the edge present and absent, eg. for Figure 4 the cases

$$f(y_1, y_2, y_3) = f_1(y_1) f_2(y_2 \mid y_1, y_3) f_3(y_3 \mid y_1)$$

or

$$= f_1(y_1)f_2(y_2|y_1)f_3(y_3|y_1) \qquad (3.2.1)$$

4. Discrete Case.

4.1 The Undirected Case.

In the case that the random variables, Y_v, $v \ \varepsilon \ V$, are discrete-valued, undirected graphical models may be described by loglinear models. With $\mathbf{Y}$ such that

$$Prob\{\mathbf{Y} = \mathbf{y}\} = \pi(\mathbf{y})$$

one may write

$$\log \pi(\mathbf{y}) = \sum_a g_a(\mathbf{y}_a)$$

where a ranges over the subsets of V. Then Y_u is independent of Y_v given the rest is equivalent to $g_a = 0$ if $\{u,v\}$ is in a , see Whittaker (1990), Edwards (1995).

The conditional independence hypothesis may again be examined by the edge exclusion deviance, that is the change in -2 *loglikelihood* when the edge is removed from the graph.

4.2 The Directed Case.

Now one would work with conditional distributions, given parents, and a joint probability function as in (2.2) and in the case of Figure 4 for factorizations like (3.2.1).

4.3 Extensions.

There are mixed discrete/continuous models, see eg. Lauritzen (1989), Whittaker (1990), Edwards (1995). In the contemporary research on this topic, given the discrete variables, the continuous variables are typically assumed jointly Gaussian.

5. Time Series.

Now random sequences and functions are under consideration.

5.1 The Undirected case.

Consider the stationary vector-valued series $\mathbf{Y}(t)$ where t is discrete or continuous. The strength of association of the u-th and v-th components may be measured by the *crosscorrelation function*

$$corr\{Y_u(t+s), Y_v(t)\} \tag{5.1.1}$$

as a function of lag s or alternately by the *coherency function*

$$R_{uv}(\lambda) = \lim_{T\to\infty} corr\{\sum_t e^{-i\lambda t} Y_u(t), \sum_t e^{-i\lambda t} Y_v(t)\} \tag{5.1.2}$$

as a function of frequency λ. As written, in (5.1.2), time is discrete and the sums are over $t = 0, \dots, T-1$. Both (5.1.1) and (5.1.2) are identically 0 in the case that the series Y_u and Y_v are statistically independent. The coherency matrix is a close analog of a correlation matrix being the correlations of the components of frequency λ in the series Y_u and Y_v, see Brillinger (1975).

In the case of a stationary Gaussian time series $\{\mathbf{Y}(t)\}$, the mutual information of the components Y_u and Y_v is

$$-\frac{1}{2\pi}\int \log\{1 - |R_{uv}(\lambda)|^2\}d\lambda$$

see Pinsker (1964), Granger and Hatanaka (1964). The coherency function (5.1.2) occurs in this last expression. The measure is 0 when $R_{uv} \equiv 0$ and ∞ when $R_{uv} \equiv 1$.

Suppressing the dependence on λ, the partial coherency of the series Y_u and Y_v, removing the linear time invariant effects of the other series, is is given by

$$R_{uv|rest} = (R_{uv} - R_{u|rest}\bar{R}_{v|rest})/\sqrt{(1-|R_{u|rest}|^2)(1-|R_{v|rest}|^2)} \tag{5.1.3}$$

with the overbar denoting complex conjugate. This is the entry in row u, column v of the inverse of the coherency matrix, $\mathbf{R}(\lambda) = [R_{uv}(\lambda)]$. It is a function of frequency, λ and so distinguished from the continuous variate case of Section 3.1. This function will be identically 0 for stationary Gaussian $\mathbf{Y}$ when Y_u and Y_v are independent given the rest. References to the concept of conditional coherency include: Tick (1963),

Granger and Hatanaka (1964), Brillinger and Hatanaka (1969,1970), Gersch (1972).

Following Granger and Hatanaka (1964), the partial mutual information of Y_u and Y_v given the rest is

$$- \frac{1}{2\pi} \int \log\{1 - |R_{uv \mid rest}(\lambda)|^2\} d\lambda$$

and one sees the appearance of the partial coherency. Coherency is an acausal concept, being identically 1 if either series involved is a linear time invariant filtered version of the other, i.e. two-sided filters are allowed.

Sometimes one may wish to consider one-sided models and corresponding likelihoods. For example consider the stationary autoregressive Gaussian case. Suppose

$$\mathbf{A}[\mathbf{Y}](t) = \mathrm{E}(t)$$

with $\mathbf{Y}$ and E r-vector-valued, with

$$\mathbf{A}[\mathbf{Y}](t) = \mathbf{a}(0)\mathbf{Y}(t) + \mathbf{a}(1)\mathbf{Y}(t-1) + \cdots + \mathbf{a}(p)\mathbf{Y}(t-p)$$

and with

$$var \ \mathrm{E}(t) = \sigma^2 \mathbf{I}, \quad cov\{\mathrm{E}(t+s),\mathrm{E}(t)\} = 0 \quad \text{for } s \neq 0$$

consider looking for 0 partial coherencies. The inverse of the spectral density matrix is

$$2\pi\overline{\mathbf{A}(\lambda)}^\tau \mathbf{A}(\lambda)/\sigma^2$$

and one sees that the entry u, v is 0 for all λ if and only if

$$\sum_{w=1}^{r} \sum_{0 \leq t+s, t \leq p} a_{wu}(t+s)a_{wv}(t) = 0$$

for all s.

The strength of connection in these two cases may be estimated by the change in deviance when model is fit both with some entries of $\mathbf{A}(.)$ identically zero and not.

Some distributions and estimates appropriate to the problems considered in this section may be found in Andersen et al. (1995).

5.2 The Directed Case.

One would set down a joint density function appropriate to the situation. In the case of Figure 4 it might be derived from

$$Y_1 \ free$$

$$Y_2(t) = \sum_{s \geq 1} a_{21}(s)Y_1(t-s) + \sum_{s \geq 1} a_{23}(s)Y_3(t-s) + \varepsilon_2(t)$$

$$Y_3(t) = \sum_{s \geq 1} a_{31}(s)Y_1(t-s) + \varepsilon_3(t)$$

with $\varepsilon_2(t)$ and $\varepsilon_3(t)$ independent of past values of Y_1, Y_2, Y_3.

5.3 The State Space Case.

The book Harvey (1989) shows the ubiquity of the state space model in economics and econometrics. Vector ARMA's and ARMAX's are particular cases of this model. So too are models with complex, nonstationary trends and behavioral econometric models. Included too is the case of time varying parameters.

One formulation is provided by

$$\mathbf{S}_t = \mathbf{A}\mathbf{S}_{t-1} + \mathbf{B}\mathbf{X}_t + \mathbf{E}_t$$
$$\mathbf{Y}_t = \mathbf{C}\mathbf{S}_t + \boldsymbol{\varepsilon}_t$$

with $\mathbf{S}_t$ the unobserved state vector, $\mathbf{X}_t$ an exogenous series and $\mathbf{Y}_t$ the endogenous. The variates $(\mathbf{E}_t, \boldsymbol{\varepsilon}_t)$ represent noise. The first equation is the transition equation, while the second is the observation.

This state space model may be usefully described by a graph, see Figure 5, and Normand and Tritcher (1992). The Kalman filter and algorithms developed by researchers in graphical modelling may be employed to predict and smooth $\mathbf{Y}_t$ here.

6. Point Processes.

The realizations of *stochastic point processes* are sets of points on the real line

$$\cdots \ < \tau_0 < \tau_1 < \tau_2 < \ \cdots$$

For example the τ's might correspond to times of: economic decisions, security sales, or bankruptcies. A point process can also be described via counting functions, specifically with I an interval one can set

$$N(I) = \#\{\tau_n \ in \ I\} \qquad N(t) = N(0,t]$$
$$dN(t) = 1 \quad if \ point \ in \ (t,t+dt]$$
$$= 0 \quad otherwise$$

References to point process analysis include: Snyder (1975), Daley and Vere-Jones (1988).

6.1 Point Process Descriptions.

Under general conditions, the *conditional intensity function* characterizes a vector-valued point process **N**. It is given by the vector-valued variate $[\mu_u(t|H_t)]$ of

$$Prob\{dN_u(t) = 1|H_t\} = \mu_u(t|H_t)dt$$

where $H_t = \{d\mathbf{N}(s), s \leq t\}$ The conditional intensity provides the likelihood and so is basic to the analysis of point process data, see Snyder (1975), Daley and Vere-Jones (1988).

6.2 The Undirected Case.

The *crossproduct density*, of the components u and v of the stationary point process **N**, provides a measure of strength of association. and is given by

$$Prob\{dN_u(t+u) = 1 \ and \ dN_v(t) = 1\}/dtdu \qquad (6.1.1)$$

The association may also be studied via the *coherency function*

$$R_{uv}(\lambda) = \lim_{T\to\infty} corr\{\sum_m e^{-i\lambda\sigma_m}, \sum_n e^{-i\lambda\tau_n}\} \qquad (6.1.2)$$

where $\sum_m$ and $\sum_n$ sum over points of the two types available in $(0,T]$. These concepts are analogs of the crosscovariance and coherency functions of time series mentioned above. There are also hybrid measures for time series and point processes.

The partial coherency of components u and v, given the rest, is again defined by (5.1.3). One can make two arguments for considering the partial coherency as a measure of conditional independence in the point process case. First, the empirical Fourier transforms appearing in (6.1.2) satisfy a Central Limit Theorem, i.e. are approximately normal and one can look to the inverse of their covariance matrix as a measure of conditional dependence. As an alternate argument, suppose that it is possible to write

$$dN_u(t) = [\sum_{j \neq u,v} \int a_j(t-s)dN_j(s)]dt + d\Gamma_u(t)$$

with Γ_u a noise process such that

$$E\{d\Gamma_u(t)|N_j, \, j \neq u,v\} = 0$$

and

$$cov\{d\Gamma_u(s),d\Gamma_v(t)\} = 0$$

then the partial coherency of Y_u and Y_v will be 0.

6.3 Directed Case.

Suppose there is a directed network with three nodes having point processes as in Figure 4. The arrows now mean that, for example, the behavior of process 2 at time t depends on the past of the other two processes. A structural model leading to associated point processes Y_1 and Y_2 in the presence of a third point process Y_3, follows. Write $Y_1 = \{\sigma_m\}$ and $Y_3 = \{\gamma_o\}$ and

$$U(t) = \sum_{\sigma_m < t} a_{21}(t - \sigma_m) + \sum_{\gamma_o < t} a_{23}(t - \gamma_o)$$

for summation functions $a_{21}(.)$ and $a_{23}(.)$ Suppose now that a Y_2 point occurs at time t when $U(t)$ upcrosses a random threshold $\theta(t)$. The question of whether Y_2 and Y_1 are independent given Y_3 in this case may be phrased as: is $a_{23}(.) \equiv 0$? Given data this question may be addressed by fitting the model both with and without $a_{23}(.)$ present.

As an approximation to the likelihood based on the conditional intensity one pro-

cedure is to replace the point processes here by 0–1 discrete time series M_t, N_t, O_t, and to estimate parameters by maximizing the log likelihood

$$\sum_t [N_t \log \Phi(U_t - \mu) + (1 - N_t) \log(1 - \Phi(U_t - \mu))]$$

where, for example, $\Phi(.)$ is the standard normal cumulative, and

$$U_t = \sum_{u=1}^{\infty} m_{t-u} M_u + \sum_{u=1}^{\infty} o_{t-u} O_u$$

and

$$\theta_t = \mu + \varepsilon_t$$

Here it may be assumed that noise values ε_t are independent identically distributed normals. The strength of the edge $3 \rightarrow 2$ may be measured by the change in deviance occurring when the function o_t is set to 0 identically.

For related models and empirical analyses, motivated by biological neuron networks see Brillinger (1991,1995).

Explanatory variables may be included, eg. one corresponding to the presence of an intervention. Set $S(t) = 1$ when the intervention is present and $= 0$ otherwise. One might assume

$$dN_u(t) = 1 \ \ \text{if} \ \ \alpha S(t) + \sum_{v \neq u} \int_0^{\infty} a_{uv}(t-s)dN_v(s) \ upcrosses \ \theta_u(t)$$

As a measure of the strength of connection one can consider the change in deviance resulting from refitting with some $a_{uv}(.) = 0$.

7. Difficulties.

There may be unmeasured variates or measurement error, see Robinson (1986). Interpretations can be fraunt with difficulties, as in the case of path analysis, see Freedman (1987). Feedback may be present, complicating the construction of the likelihood function.

8. Discussion and Summary.

In distinction with the ordinary random variable case, the time series and point process cases lead to the consideration of parameters that are functions, functions of frequency or lag.

Graphical representations have been found to be valuable in the sense that they force one to think hard about the characters of the dependencies involved. Such considerations are basic to economics and econometrics, see Hendry and Richards (1983). Econometricians are often lead to think about the issue of causality, see eg. Geweke (1984), and the current work on causality and graphs, see eg. Pearl (1995) and its discussion, has pertinent specific techniques to offer.

The referee made the basic point that temporal systems can be time-varying, the graph may change (as an example I mention that nodes may appear or dissappear) and a more sophisticated structure is needed. I agree and thank the Referee for this and other comments.

Experimental systems can provide a testbed for econometric techniques. The techniques presented here have been validated, to an extent, by experimental data from neurophysiology, see Rosenberg et al (1989), Brillinger (1991,1995).

References

Andersen, H.H., Hojbjerre, Sorensen, D. and Eriksen, P.S. 1995. Linear and Graphical Models. New York, Springer.

Brémaud, P. 1981. Point Processes and Queues. New York, Springer.

Brillinger, D.R. 1975. Time Series: Data Analysis and Theory. New York, Holt Rinehart.

Brillinger, D.R. 1991. "Nerve cell spike train data analysis: a progression of technique." J. Amer. Statist. Assoc. 87:260-271.

Brillinger, D.R. 1995. "Networks of time series and point processes" In Morgenthaller,

S., John W. Tukey Festschrift. New York, Springer. To appear.

Brillinger, D.R. and Hatanaka, M. 1969. "An harmonic analysis of nonstationary multivariate economic processes." Econometrica 131-141.

Brillinger, D.R. and Hatanaka, M. 1970. "A permanent income hypothesis relating to the aggregate demand for money (an application of spectral and moving spectral analysis)." Economic Studies Quarterly 21:44-71.

Cox, D.R. and Wermuth, N. 1993. "Linear dependencies represented by chain graphs." Statistical Science, 8: 204-218.

Daley, D.J. and Vere-Jones, D. 1988. An Introduction to the Theory of Point Processes. New York, Springer.

Dempster, A.P. 1972. "Covariance selection." Biometrics 28: 157-175.

Edwards, D. 1995. Introduction to Graphical Modelling. New York, Springer.

Freedman, D.A. 1987. "As others see us: A case study in path analysis". J. Educational Stat. 12:101-128.

Geiger, D. and Pearl, J. 1993. "Logical and algorithmic properties of conditional independence and graphical models." Ann. Statist. 21:2001-2021.

Gersch, W. 1972. "Causality or driving in electrophysiological signal analysis." Math. Biosci. 14:177-196.

Geweke, J. (1984). "Inference and causality in economic time series models." Z. Griliches and M.D. Intriligator, Handbook of Econometrics Vol. II. Amsterdam, Elsevier.

Glymour, C., Scheines, R., Spirtes, P. and Kelly, K. 1987. Discovering Causal Structure. New York, Academic.

Granger, C.W.J. and Hatanaka, M. 1964. Spectral Analysis of Economic Time Series. Princeton, Princeton Press.

Granger, C.W.J. 1980. "On the synthesis of time series and econometric models." In Brillinger, D.R. and Tiao, G.C., Directions in Time Series. Hayward, Institute of Mathematical Statistics.

Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge, Cambridge University Press.

Hendry, D.F. and Richard, J-F. 1983. "The econometric analysis of economic time series." Int. Statist. Review 51:111-148.

Jacquez, J.A. 1972. Compartmental Analysis in Biology and Medicine. Amsterdam, Elsevier.

Kanter, M. 1979. "Lower bounds for nonlinear prediction error in moving average processes." Ann. Prob. 7:128-138.

Kiiveri, H. and Speed, T.P. 1982. "Structural analysis of multivariate data: a review". In Leinhardt, S., Sociological Methodology. San Francisco, Jossey-Bass.

Lauritzen, S.J. 1989. "Mixed graphical association models." Scand. J. Statist. 16:273-306.

Lauritzen, S.L., Dawid, A.P., Larsen, B.N. and Leimer, H-G. 1990. "Independence properties of directed Markov fields." Networks, 20:491-505.

Normand, S-L. and Trichler, D. "Parameter updating in a Bayes network." J. American Statistical Assoc. 87:1109-1115.

Parzen, E. 1983. "Time series model identification by estimating information." In Karlin, S., Amemiya, T. and Goodman, L.A., Studies in Econometrics, Time Series and Multivariate Statistics. New York, Academic.

Pearl, J. "Causal diagrams for empirical research." Biometrika 82:669-688.

Philips, A.W. 1950. "Mechanical models in economic dynamics." Economica, 17:282-305.

Pinsker, M.S. 1964. Information and Information Stability of Random Variables and Processes. San Francisco, Holden-Day.

Robinson, P.M. 1986. "On the errors-in-variables problem for time series." J. Mult. Analysis 19:240-250.

Rosenberg, J.R., Amjad, A.M., Breeze, P., Brillinger, D.R. and Haliday, D.M. 1989. "The Fourier approach to the identification of functional coupling between neuronal spike trains." Prog. Biophys. Molec. Biol. 53:1-31.

Smith, J.Q. 1989. "Influence diagrams for statistical modelling." Ann. Statist. 17:654-672.

Snyder, D.L. 1975. Random Point Processes. New York, Wiley.

Swade, D. 1995. "The Philips economic computer." Resurection, No. 12: 11-18.

Tick, L.J. 1963. "Conditional spectra, linear systems and coherency." In Rosenblatt, M., Time Series Analysis. New York, Wiley.

Tinbergen, J. 1939. Business Cycles in the United States of America 1919-1932. Geneva, League of Nations.

Wahba, G. (1971). "Some tests of independence of stationary multivariate time series." J. Royal Statist. Soc. B 33:153-166.

Wermuth, N. 1992. "On block-recursive linear regression equations." REBRAPE, 6:1-56.

Wermuth, N. and Lauritzen, S.L. 1990. "On substantive research hypotheses, conditional independence graphs and graphical chain models." J. Royal Statist. Soc. B 52: 21-72.

Whittaker, J. 1990. Graphical Models in Applied Mathematical Multivariate Statistics. New York, Wiley.

Wold, H. 1956. "Causal inference from observational data." J. Roy. Statist. Soc. A 119:28-61.

Legends

Figure 1. Philips Simple Model. C: consumption, E: total expenditure, I: investment
expenditure, M: surplus balances, M': transfer balances, S: savings, Y: income.

Figure 2. Example of an undirected graph, with 4 nodes.

Figure 3. Examples of a directed graph with 4 nodes and 2 directed edges.

Figure 4. Example of a directed graph with 3 nodes and 3 directed edges.

Figure 5. A directed graph representing the state space model with the exogenous
variable $\mathbf{X}_t$, the state variable $\mathbf{S}_t$ and the endogenous variable $\mathbf{Y}_t$.