MARKOV PROPERTIES OF NONRECURSIVE CAUSAL MODELS

By J. T. A. Koster

Erasmus University Rotterdam

This paper aims to solve an often noted incompatibility between graphical chain models which elucidate the conditional independence structure of a set of random variables and simultaneous equations systems which focus on direct linear interactions and correlations between random variables. Various authors have argued that the incompatibility arises mainly from the fact that in a simultaneous equations system (e.g., a LISREL model) reciprocal causality is possible whereas this is not so in the case of graphical chain models. In this article it is shown that this view is not correct. In fact, the definition of the Markov property embodied in a graph can be generalized to a wider class of graphs which includes certain nonrecursive graphs. The resulting class of reciprocal graph probability models strictly includes the class of chain graph probability models. The class of lattice conditional independence probability models is also strictly included. It is shown that the resulting methodology is directly applicable to quite general simultaneous equations systems that are subject to mild restrictions only. Provided some adjustments are made, general simultaneous equations systems can be handled as well. In all cases, consistency with the LISREL methodology is maintained.

1. Introduction. Recent years have shown the development of the theory of graphical models: statistical models of which the Markov properties (i.e., the conditional independence structure of a set of random variables) are accurately portrayed by a certain graph [Frydenberg (1990a), Lauritzen and Wermuth (1989) and Whittaker (1990)]. These models generalize the older theory of recursive path models [Wright (1934) and Wermuth (1980)], the covariance selection models for continuous variables [Dempster (1972)] and the graphical log-linear models for discrete variables [Lauritzen (1979) and Darroch, Lauritzen and Speed (1980)]. In Kiiveri, Speed and Carlin (1984) the factorization of the joint normal density of a set of random variables satisfying a recursive system of linear equations (e.g., a LISREL model) was studied, and it was shown that some of the Markov properties of such a system can be read off the associated path diagram, provided certain assumptions on the covariance matrices of the exogenous and error variables are met. Assuming uncorrelated exogenous and error variables, Lauritzen, Dawid, Larsen and Leimer (1990) have strengthened the results of Kiiveri, Speed

and Carlin (1984) such that more Markov properties can be read off the path diagram. In summary, the theory of graphical models developed thus far covers undirected graphs (considering only symmetric interactions between variables), directed graphs (considering only causal relationships between variables) and chain graphs (= block-recursive models). These models exclude causal cycles or bidirectional influences.

The motivation for the present article is the circumstance, expressed by various authors [e.g., Kiiveri and Speed (1982), page 239, Lauritzen (1989), page 302, and Whittaker (1990), pages 72, 302–303], that the theory of graphical chain models generates inconsistencies when applied to models for general simultaneous equations (SE) systems, that is, systems in which causal cycles between variables occur. Therefore, one cannot, for example, read the Markov properties of a LISREL model from the associated LISREL path diagram in case the system of linear equations is nonrecursive.

The article roughly consists of two parts. In the first part (Sections 2 and 3) a class of graphs is introduced (i.e., the reciprocal graphs) which is much wider than the class of chain graphs, and the definition of the Markov property is extended to this class of graphs. In this way a broad set of nonrecursive systems can be modeled which strictly contains the class of probability models determined by chain graphs. It is also shown that the "Gibbs = Markov" equivalence [cf. Speed (1979)], a key result for statistical applications, holds for reciprocal graphs. In the second part (Sections 4 to 6) we show that for a restricted class of simultaneous equations systems this extension is consistent with the LISREL methodology in the following sense: all conditional independence statements that can be derived from the reciprocal graph that is associated with the simultaneous linear equations system are valid for all normal probability distributions satisfying the simultaneous equations system. As to the converse, it is shown that, if a conditional independence statement is valid for all normal probability distributions satisfying the simultaneous equations system, then the statement can be read off the graph. It is also explained how general simultaneous equations systems, which may violate the restrictive conditions referred to above, can be dealt with.

The structure of the paper is as follows. In Section 2 the necessary graph theoretical terminology is presented and the notion of a reciprocal graph is introduced. Also, basic results on finite distributive (set) lattices are presented. In Section 3 the Markov property for reciprocal graphs is defined. This definition generalizes the corresponding definition for chain graphs as given by Frydenberg (1990a). The main result of this section is the "Gibbs = Markov" equivalence expressed in Theorem 3.4. We also clarify the relation between graphical probability models and the lattice conditional independence (CI) probability models introduced by Andersson and Perlman (1993). In Section 4 we consider simultaneous equations systems (e.g., LISREL models) and discuss some of the inconsistencies which may arise when a LISREL path diagram is misinterpreted as a reciprocal graph embodying certain Markov properties. A solution to the problems is proposed in the form

of two extra assumptions concerning the zero structure of the covariance matrices of the exogenous and error variables. In Section 5 the main theorem is stated and proved. It is also asserted there that the class of normal probability distributions satisfying a SE system that meets certain conditions is Markov perfect. A direct proof of a restricted version of this assertion is given. An illustrative example, showing that the class of probability models determined by a reciprocal graph is strictly larger than the class of models determined by a chain graph, ends this section. In Section 6 we briefly discuss how to handle simultaneous equations systems that breach the assumptions concerning the zero structure of the covariance matrices of the exogenous and error variables. This discussion is based on recent results of Spirtes (1995) and Spirtes, Richardson, Meek, Scheines and Glymour (1996) and allows us to fully establish the Markov perfectness claim. The upshot of Sections 5 and 6 is that neither nonrecursiveness nor correlated exogenous or error variables pose any problem to reading the Markov properties off a LISREL path diagram. Some open questions are discussed in Section 7.

2. Reciprocal graphs. In this section we will define the type of graph we intend to study. Since one of our objectives is to generalize certain results of Frydenberg (1990a), we will mainly conform to the notation and definitions of that paper. In the process, concepts of Andersson and Perlman (1993) pertaining to *lattice CI models* will also be introduced. In cases where our concepts are defined differently from Frydenberg (1990a), this is done solely for the purpose of using them in a wider context. In such cases the reader is invited to verify that our definitions are equivalent to the definitions of Frydenberg when applied to the more narrow context.

A *graph* is a pair $G = (V, E)$, where $V = V(G)$ is a finite set of vertices and $E = E(G)$ is a set of directed or undirected edges. Undirected edges are members of the class $\mathscr{P}_2(V) = \{\{\alpha, \beta\} | \alpha \in V, \beta \in V \text{ and } \alpha \neq \beta\}$. Directed edges belong to $E^*(V) = \{(\alpha, \beta) | \alpha \in V, \beta \in V \text{ and } \alpha \neq \beta\}$. In drawing a picture of a graph, undirected edges are represented by a line segment (e.g., $\alpha - \beta$ when $\{\alpha, \beta\} \in E$), whereas directed edges are denoted by an arrow pointing from the first member of the ordered pair to the second member [e.g., $\alpha \rightarrow \beta$ when $(\alpha, \beta) \in E$]. If both (α, β) and (β, α) belong to E, two arrows are drawn (e.g., $\alpha \rightleftarrows \beta$). Notice that this differs from Frydenberg (1990a) where the condition $\{(\alpha, \beta), (\beta, \alpha)\} \subseteq E$ denotes that α and β are connected by an undirected edge. Here we make a clear distinction between $\alpha \rightleftarrows \beta$ and $\alpha - \beta$, and it is the latter, *not the first*, which is equivalent to Frydenberg's condition $\{(\alpha, \beta), (\beta, \alpha)\} \subseteq E$. A subset of vertices is called *complete* if each pair of its vertices is connected by an edge. A maximal complete subset of vertices is called a *clique*.

Suppose $G = (V, E)$ and $H = (W, F)$ are graphs. H is called a *subgraph* of G if $W \subseteq V$ and $F \subseteq E$. We also define $G \cup H = (V \cup W, E \cup F)$ and $G \cap H = (V \cap W, E \cap F)$. If a is a subset of vertices, the *subgraph induced by* a is $G_a = (a, E_a)$, where $E_a = E \cap (\mathscr{P}_2(a) \cup E^*(a))$. An ordered $(n + 1)$-tuple $(\alpha_0, \ldots, \alpha_n)$ of vertices that are distinct except possibly $\alpha_0 = \alpha_n$ is called a

path of length n from α_0 *to* α_n if $\{\alpha_{i-1}, \alpha_i\} \in E$ or $(\alpha_{i-1}, \alpha_i) \in E$ for $i = 1, \ldots, n$. By $\mathrm{an}(\alpha)$ we denote the set $\{\alpha\} \cup \{\beta \mid \beta \in V$, and there exists a path from β to $\alpha\}$. Clearly, if $\beta \in \mathrm{an}(\alpha)$, then $\mathrm{an}(\beta) \subseteq \mathrm{an}(\alpha)$. A path is called *undirected* if $\{\alpha_{i-1}, \alpha_i\} \in E$ for $i = 1, \ldots, n$, otherwise it is called *directed*. Suppose $\alpha, \beta \in V$. We will write $\alpha \equiv \beta$ if $\alpha = \beta$ or if there exists an undirected path from α to β. We write $\alpha \approx \beta$ if $\alpha = \beta$ or if there exist paths from α to β and from β to α. In other words, $\alpha \approx \beta$ if and only if $\beta \in \mathrm{an}(\alpha)$ and $\alpha \in \mathrm{an}(\beta)$. Or, equivalently, $\alpha \approx \beta$ if and only if $\mathrm{an}(\alpha) = \mathrm{an}(\beta)$. It is easily seen that $\equiv$ and $\approx$ define equivalence relations on the set of vertices. The induced set of $\equiv$-equivalence classes is denoted by $U(G)$; the $\equiv$-equivalence class of vertex α is denoted by $u(\alpha)$ and is called the (*undirected*) *path component of* α. The induced set of $\approx$-equivalence classes is denoted by $C(G)$; the $\approx$-equivalence class of vertex α is denoted by $c(\alpha)$ and is called the *cycle component of* α. Notice that $u(\alpha) \subseteq c(\alpha) \subseteq \mathrm{an}(\alpha)$ for all $\alpha \in V$.

For a vertex $\alpha \in V$, the *boundary of* α is defined as the set $\mathrm{bd}(\alpha) = \{\beta \mid \beta \in V, \{\beta, \alpha\} \in E$ or $(\beta, \alpha) \in E\}$. Since the boundary of α consists of all vertices β for which there is a path of length 1 to α, $\mathrm{bd}(\alpha) \subseteq \mathrm{an}(\alpha)$ for all $\alpha \in V$. For a subset $a \subseteq V$, the *boundary of* a is defined by $\mathrm{bd}(a) = (\bigcup_{\alpha \in a} \mathrm{bd}(\alpha)) \setminus a$; the *closure of* a is defined by $\mathrm{cl}(a) = a \cup \mathrm{bd}(a)$. In what follows an important role is played by subsets which have an empty boundary. These sets are called *anterior sets*, so $a \subseteq V$ is called an anterior set of the graph G if $\mathrm{bd}(a) = \varnothing$. This of course means that, for all $\alpha \in a$, $\mathrm{bd}(\alpha) \subseteq a$. Note that $\varnothing$ and V are anterior sets.

LEMMA 2.1. *Let* $G = (V, E)$ *be a graph and suppose* $a \subseteq V$. *Then* a *is an anterior set if and only if, for all* $\alpha \in a$, $\mathrm{an}(\alpha) \subseteq a$.

PROOF. Suppose a is an anterior set, $\alpha \in a$ and $\beta \in \mathrm{an}(\alpha)$. Then there exists a path, say $(\alpha_0, \ldots, \alpha_n)$, from $\beta = \alpha_0$ to $\alpha = \alpha_n$. Since either $\{\alpha_{n-1}, \alpha\} \in E$ or $(\alpha_{n-1}, \alpha) \in E$, $\alpha_{n-1} \in \mathrm{bd}(\alpha)$, hence $\alpha_{n-1} \in a$ and $\beta \in \mathrm{an}(\alpha_{n-1})$. Continuing in this way, we find after $n - 1$ further steps that $\alpha_0 \in a$. The other way around is trivial. $\square$

According to our next lemma, an anterior subset of an anterior set is still an anterior set in the original graph.

LEMMA 2.2. *Let* $G = (V, E)$ *be a graph and suppose* $a \subseteq V$ *is an anterior set in* G. *Suppose* $b \subseteq a$ *is an anterior set in the induced subgraph* $G_a = (a, E_a)$. *Then* b *is also an anterior set in* G.

PROOF. We must show that b contains the boundary (in G) of each of its points. Suppose $\beta \in b$, $\alpha \in \mathrm{bd}(\beta)$, where the boundary is taken w.r.t. the graph G. Since $b \subseteq a$ and a is an anterior set in G, $\alpha \in a$. But then the particular edge—be it $\{\alpha, \beta\}$ or (α, β)—is a member of E_a, hence $\alpha \in \mathrm{bd}(\beta)$, where now the boundary is taken w.r.t. the subgraph G_a. Since b is an anterior set in G_a, it follows that $\alpha \in b$. $\square$

It is easy to see that the class of all anterior sets of the graph $G = (V, E)$ is closed under set union and intersection (recall that $\varnothing$ and V are anterior sets). This means that the anterior sets of G constitute a finite distributive lattice under the join and meet operations $\cup$ and $\cap$, respectively. For future reference we state this important fact as a proposition.

PROPOSITION 2.3. *Let $G = (V, E)$ be a graph. The class of anterior sets of G constitutes a finite distributive lattice with $\cup$ and $\cap$ as join and meet operations.*

In the sequel some general properties of (set) lattices will be used. For the sake of easy reference these properties will be summarized below. For proofs and more information the reader is referred to Andersson and Perlman (1993).

Suppose $A \subseteq \mathscr{P}(V)$ is closed under $\cup$ and $\cap$ and satisfies $\{\varnothing, V\} \subseteq A$, so A is finite distributive lattice with $\cup$ and $\cap$ as join and meet operations. For $a \in A$, define $\langle a \rangle = \bigcup\{b \in A | b \subset a\}$ and $[a] = a \setminus \langle a \rangle$. Notice that $\langle a \rangle \in A$. The set $J(A) = \{a \in A \| [a] \neq \varnothing\} = \{a \in A | \langle a \rangle \subset a\}$ is called the class of *join-irreducible* elements of the lattice A. There always exists a *never-decreasing* listing of the members of $J(A)$, that is, a listing $a_1, \ldots, a_q$ such that $i < j \Rightarrow a_j \not\subseteq a_i$; $q = |J(A)|$.

For $b \in A$, define $A_b = \{a \in A | a \subseteq b\}$, the lattice induced on b by A. Important facts concerning lattices and their join-irreducible elements are collected in the following proposition.

PROPOSITION 2.4. *Suppose $b, c \in A$. Then, denoting disjoint union by $\dot\cup$:*

(i) $b = \bigcup J(A_b)$.

(ii) $J(A_b) = J(A) \cap A_b$.

(iii) $J(A_{b \cap c}) = J(A_b) \cap J(A_c)$.

(iv) $J(A_{b \cup c}) = J(A_b) \cup J(A_c)$.

(v) $b = \dot\bigcup\{[a] | a \in J(A_b)\}$.

(vi) $V \setminus b = \dot\bigcup\{[a] | a \in J(A) \setminus J(A_b)\}$.

(vii) If $a_1, \ldots, a_q$ is a never-decreasing listing of the members of $J(A)$, then, for $1 \leq i \leq q$, $J(A_{a_i}) \subseteq \{a_j | j \leq i\}$, $\bigcup_{j \leq i}[a_j] = (\bigcup_{j \leq i} a_j) \in A$ and $b \setminus (\bigcup_{j \geq i}[a_j]) \in A$.

(viii) There exists a never-decreasing listing $a_1, \ldots, a_q$ of the members of $J(A)$ such that if $[a_i] \subseteq b$ and $[a_j] \not\subseteq b$, then $i < j$.

PROOF. For (i) to (v) see Andersson and Perlman (1993). Relation (vi) follows from (v) and the fact that $J(A) = J(A_V)$. Notice that it follows from (v) and (vi) that $a \in J(A)$, $b \in A$, $[a] \subseteq b$ implies $a \subseteq b$. To see the first statement of (vii), note that if $j > i$, then $a_j \not\subseteq a_i$, hence $a_j \notin J(A_{a_i})$. The second statement follows from the first and (v) and the fact that A is closed under set union. The third statement follows from $b \setminus (\bigcup_{j \geq i}[a_j]) = b \cap (\bigcup_{j < i}[a_j]) = b \cap (\bigcup_{j < i} a_j)$. To show (viii), let $a_1, \ldots, a_q$ be a never-decreasing listing of $J(A)$ and assume $b \neq V$ ($b = V$ is trivial). Using (v) and (vi), we

have $\{i | 1 \leq i \leq q, [a_i] \subseteq b\} = \{i_1, \ldots, i_p\}$, where $i_1 < \cdots < i_p$, and, similarly, $\{i | 1 \leq i \leq q, [a_i] \nsubseteq b\} = \{i_{p+1}, \ldots, i_q\}$, where $i_{p+1} < \cdots < i_q$. Now, $a_{i_1}, \ldots, a_{i_p}$ and $a_{i_{p+1}}, \ldots, a_{i_q}$ are both never-decreasing sequences. Furthermore, if $1 \leq r \leq p < s \leq q$, then $[a_{i_r}] \subseteq b$, hence $a_{i_r} \subseteq b$, so $a_{i_s} \nsubseteq a_{i_r}$ since otherwise $a_{i_s} \subseteq b$, contradicting $[a_{i_s}] \nsubseteq b$. So the listing $a_{i_1}, \ldots, a_{i_p}, a_{i_{p+1}}, \ldots, a_{i_q}$ is never-decreasing and satisfies (viii). $\square$

The lattice of all anterior sets of the graph $G = (V, E)$ will be denoted by $A(G)$. It will often be necessary to identify the elements of $A(G)$. In general, this can most readily be done by first identifying the path or cycle components in G and then forming unions of these components, always including with each component all components from which an arrow emanates pointing toward it. The lattice $A(G)$ satisfies some additional properties.

PROPOSITION 2.5.   *Suppose $G = (V, E)$ is a graph and $a \in A(G)$. Then*:

(i) $(A(G))_a = A(G_a)$;
(ii) $\alpha \in [a] \Leftrightarrow \mathrm{an}(\alpha) = a$;
(iii) $\alpha \in [a] \Leftrightarrow c(\alpha) = [a]$;
(iv) $\mathrm{bd}([a]) = \{\beta \in V \mid \beta \in \langle a \rangle; (\beta, \alpha) \in E \text{ for some } \alpha \in [a]\}$;
(v) $\alpha \in \langle a \rangle \Leftrightarrow \mathrm{an}(\alpha) \subseteq \langle a \rangle$.

PROOF.   Note that (iii), (iv) and the necessity implication of (ii) are trivially satisfied if $a \in A(G) \setminus J(A(G))$.

(i) $(A(G))_a \subseteq A(G_a)$ is trivial. The other way around follows from Lemma 2.2.

(ii) Suppose $\alpha \in [a]$. Note that $\mathrm{an}(\alpha) \subseteq a$, $\mathrm{an}(\alpha) \in A(G)$, so if $\mathrm{an}(\alpha) \neq a$, then $\mathrm{an}(\alpha) \subseteq \langle a \rangle$, contradicting $\alpha \in [a]$. Now assume $\mathrm{an}(\alpha) = a$. It follows from (i) and Proposition 2.4(v) that $\alpha \in [b]$ for some $b \in J(A(G_a))$. From what we have already proved, it follows that $\mathrm{an}(\alpha) = b$ (in G_a and hence also in G, hence $a = b$ and consequently $\alpha \in [a]$.

(iii) Sufficiency is trivial since $\alpha \in c(\alpha)$. For necessity, let $\alpha \in [a]$. From the definition of $c(\alpha)$ and using (ii) twice, we have $c(\alpha) = \{\beta \in V \mid \mathrm{an}(\beta) = \mathrm{an}(\alpha)\} = \{\beta \in V \mid \mathrm{an}(\beta) = a\} = [a]$.

(iv) Clearly, $\mathrm{bd}([a]) \subseteq \mathrm{an}([a]) \subseteq a$, so $\mathrm{bd}([a]) \subseteq a \setminus [a] = \langle a \rangle$. Suppose $\beta \in \mathrm{bd}([a])$, say $\beta \in \mathrm{bd}(\alpha)$ for some $\alpha \in [a]$. Using (iii), $\{\alpha, \beta\} \notin E$, otherwise $\beta \in u(\alpha) \subseteq c(\alpha) = [a]$. So $(\beta, \alpha) \in E$. The inclusion $\supseteq$ is obvious.

(v) This follows from the definition of $\mathrm{an}(\alpha)$ and the fact that $\langle a \rangle \in A(G)$.
$\square$

COROLLARY 2.6.   (i) $C(G) = \{[a] \mid a \in J(A(G))\}$.
(ii) $J(A(G)) = \{\mathrm{an}(\alpha) \mid \alpha \in V\}$.

So the sets $[a], a \in J(A(G))$ can be referred to as cycle components.

If $c \subseteq V$, then the smallest anterior set containing c is denoted by $\mathrm{an}(c)$. Proposition 2.3 shows that this set is well defined. No confusion can arise in case $c = \{\alpha\}$, since $\mathrm{an}(\{\alpha\}) = \mathrm{an}(\alpha)$ indeed. Using Proposition 2.3, it will also be clear that the following result holds.

LEMMA 2.7.   *If a and b are subsets of V, then* $\mathrm{an}(a \cup b) = \mathrm{an}(a) \cup \mathrm{an}(b)$.

Our definition of a graph allows for the possibility that, for two vertices α and β, each of $\{\alpha, \beta\}$ and (α, β) and (β, α) are edges. In fact, this is a bit too much, since for such general graphs difficulties arise in properly defining the moral graph. For this reason we will restrict ourselves to a certain type of graph. We will define a *reciprocal graph* as a graph $G = (V, E)$ that satisfies the condition $E \cap E^*(u) = \varnothing$ for all $u \in U(G)$. This means that there are no directed edges between vertices which belong to the same undirected path component. Hence arrows can only run between vertices from different path components. In particular, we do not allow both $\{\alpha, \beta\} \in E$ and $(\alpha, \beta) \in E$. From now on, all graphs considered are understood to be reciprocal graphs, unless stated otherwise explicitly. In Example 1 below, (a), (b), (c) and (d) are reciprocal graphs, whereas (e) is not reciprocal.

A graph is called *undirected* if it does not contain directed edges: $E \cap E^*(V) = \varnothing$. An important notion for undirected graphs is the concept of *graph separation*. If a, b and c are disjoint subsets of V, we say that a and b are separated by c if each path from a vertex $\alpha \in a$ to a vertex $\beta \in b$ contains a vertex of c. The condition is understood to be trivially satisfied if $a = \varnothing$ or $b = \varnothing$.

A graph is called a *chain graph* if it does not contain any *directed cycles*. A directed cycle is an ordered $(n + 1)$-tuple $(\alpha_0, \ldots, \alpha_n)$ of vertices that are distinct except for $\alpha_0 = \alpha_n$, such that for $i = 1, \ldots, n$ either $\{\alpha_{i-1}, \alpha_i\} \in E$ or $(\alpha_{i-1}, \alpha_i) \in E$ and the latter occurs at least once. A chain graph can equivalently be defined as a reciprocal graph G for which $U(G) = C(G)$. Notice that in a chain graph the final cycle component $[a_q]$ in a never-decreasing listing of $J(A(G))$ is called a terminal chain component by Frydenberg (1990a). Also note that a reciprocal graph cannot contain directed cycles which have exactly one arrow (besides undirected line segments), but it may very well contain directed cycles with two or more arrows (the simplest example of this being $\alpha \rightleftarrows \beta$).

Chain graphs and their Markov properties were discussed thoroughly by Frydenberg (1990a). Although in the present article chain graphs are introduced slightly differently (i.e., as a subclass of the more general class of reciprocal graphs), a little reflection shows that they are the same mathematical objects. [Recall that Frydenberg's condition $\{(\alpha, \beta), (\beta, \alpha)\} \subseteq E$ corresponds to our condition $\{\alpha, \beta\} \in E$.] Similar remarks hold for the way we defined the concept of an "anterior set." It generalizes Frydenberg's anterior sets—defined by him exclusively in the context of chain graphs—to the context of (reciprocal) graphs, but in case the graph is a chain graph both definitions coincide.

Let $G = (V, E)$ be a reciprocal graph. The *underlying undirected graph* is defined by $G^u = (V, E^u)$, where $E^u = \{\{\alpha, \beta\}|\{\alpha, \beta\} \in E$ or $(\alpha, \beta) \in E\}$. This means that all arrows are replaced by undirected edges (at most one edge between two vertices, though). The *moral graph of G* is defined by $G^m = (V, E^m)$, where $E^m = E^u \cup (\bigcup_{u \in U(G)} \mathscr{P}_2(\text{bd}(u)))$, that is, the underlying undirected graph, where the boundary of each path component of G is made complete by means of undirected edges. Note that G^m is an undirected graph.

Some of the concepts introduced so far will be illustrated by the following example.

EXAMPLE 1.   (a) $G = (\{1, 2, 3, 4\}, \{(1, 2), (2, 3), (3, 4), (4, 1)\})$. A picture of G would look like

$$
\begin{array}{ccc}
1 & \to & 2 \\
\uparrow & & \downarrow \\
4 & \leftarrow & 3
\end{array}
$$

The path components of G are $\{\alpha\}$, $\alpha = 1, 2, 3, 4$, so G is clearly reciprocal. The anterior sets of G are $\varnothing$ and $\{1, 2, 3, 4\}$; the latter set is the only join-irreducible set and hence also the only cycle component. The moral graph G^m is just

$$
\begin{array}{ccc}
1 & — & 2 \\
| & & | \\
4 & — & 3
\end{array}
$$

(b) $G = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{3, 4\}, (2, 3), (4, 1), (1, 4)\})$. Note that the path components are $\{1, 2\}$ and $\{3, 4\}$; the anterior sets are $\varnothing$ and $\{1, 2, 3, 4\}$. Again, $\{1, 2, 3, 4\}$ is the only join-irreducible set and also the only cycle component.

$$
\begin{array}{ccc}
1 & — & 2 \\
\uparrow\downarrow & & \downarrow \\
4 & — & 3
\end{array}
$$

The moral graph G^m is the same as in example (a).

(c) $G = (\{1, 2, 3, 4\}, \{\{1, 2\}, (2, 3), (4, 1), (4, 3), (1, 4)\})$. The path components are $\{1, 2\}$, $\{3\}$ and $\{4\}$; the anterior sets are $\varnothing$, $\{1, 2, 4\}$ and $\{1, 2, 3, 4\}$. The latter two sets are also join-irreducible, so the cycle components of this graph are $[\{1, 2, 4\}] = \{1, 2, 4\}$ and $[\{1, 2, 3, 4\}] = \{3\}$.

$$
\begin{array}{ccc}
1 & — & 2 \\
\uparrow\downarrow & & \downarrow \\
4 & \to & 3
\end{array}
$$

The moral graph G^m replaces all arrows by undirected edges and adds the edge $\{2, 4\}$ since $\text{bd}(3) = \{2, 4\}$ is to be made complete. Hence G^m becomes

$$
\begin{array}{ccc}
1 & — & 2 \\
| & \diagup & | \\
4 & — & 3
\end{array}
$$

(d) $G = (\{1, 2, 3, 4\}, \{(2, 1), (2, 3), (4, 1), (3, 4), (1, 4)\})$. The path components are $\{\alpha\}$, $\alpha = 1, 2, 3, 4$; the anterior sets are $\varnothing$, $\{2\}$, $\{2, 3\}$ and $\{1, 2, 3, 4\}$; the

join-irreducible sets are $\{2\}$, $\{2, 3\}$ and $\{1, 2, 3, 4\}$, leading to the cycle components $\{2\}$, $\{3\}$ and $\{1, 4\}$.

$$
\begin{array}{ccc}
1 & \leftarrow & 2 \\
\Updownarrow & & \downarrow \\
4 & \leftarrow & 3
\end{array}
$$

The moral graph G^m again replaces all arrows by undirected edges and adds edges to make boundaries of the path components complete. Hence G^m is

$$
\begin{array}{ccc}
1 & \!\!-\!\! & 2 \\
| & \times & | \\
4 & \!\!-\!\! & 3
\end{array}
$$

(e) Finally, consider the graph $G = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{2, 3\}, \{3, 4\}, (4, 1)\})$. A picture of G looks like

$$
\begin{array}{ccc}
1 & \!\!-\!\! & 2 \\
\uparrow & & | \\
4 & \!\!-\!\! & 3
\end{array}
$$

Since $\{1, 2, 3, 4\}$ is a path component, the arrow $4 \rightarrow 1$ runs between vertices of the same path component. Hence G is not reciprocal.

In the sequel, the operation of inducing a subgraph and considering its moral graph will almost exclusively take place w.r.t. anterior sets. The following proposition may sometimes be useful. Its simple proof is left to the reader.

PROPOSITION 2.8. *Suppose a and b are anterior sets of the reciprocal graph G. Then $G_{a \cup b} = G_a \cup G_b$ and $(G_{a \cup b})^m = (G_a)^m \cup (G_b)^m$.*

We end this section with a result concerning graph separation.

PROPOSITION 2.9. *Suppose $G = (V, E)$ is a reciprocal graph. Let $a_1, \ldots, a_q$ be a never-decreasing listing of the members of $J(A(G))$. For $1 \le i \le q$, let b_i denote $\bigcup_{j \le i} a_j = \bigcup_{j \le i} [a_j]$. Then $[a_i]$ is separated from $b_i \setminus \mathrm{cl}([a_i])$ by $\mathrm{bd}([a_i])$ in $(G_{b_i})^m$. Also, $[a_i]$ is separated from $a_i \setminus \mathrm{cl}([a_i])$ by $\mathrm{bd}([a_i])$ in $(G_{a_i})^m$. [Here $\mathrm{bd}([a_i])$ and $\mathrm{cl}([a_i])$ are to be taken w.r.t. G, or, which is the same, w.r.t. one of G_{b_i}, $(G_{b_i})^m$, G_{a_i} or $(G_{a_i})^m$.]*

PROOF. Note that $b_i = a_i \cup b_{i-1}$, $b_0 := \varnothing$, so the previous proposition implies that the second statement follows from the first statement. To see the first statement, since $[a_i] \cup b_i \setminus \mathrm{cl}([a_i]) \cup \mathrm{bd}([a_i]) = b_i$, it suffices to show that $\{\alpha, \beta\}$ is not an edge in $(G_{b_i})^m$ whenever $\alpha \in [a_i]$ and $\beta \in b_i \setminus \mathrm{cl}([a_i])$. Clearly, $\{\alpha, \beta\} \notin E$ and $(\beta, \alpha) \notin E$ since otherwise $\beta \in \mathrm{bd}([a_i])$. Suppose $(\alpha, \beta) \in E$, say $\beta \in [a_k]$ for some $k < i$. Then, using Proposition 2.5 (ii) and Proposition 2.4(vii), $\alpha \in \mathrm{an}(\beta) = a_k \subseteq \bigcup\{[a_j] | j \le k\}$, contradicting $\alpha \in [a_i]$. Finally, suppose $\{\alpha, \beta\} \subseteq \mathrm{bd}(u(\gamma))$ for some $\gamma \in b_i$. If $\gamma \in [a_i]$, then $\beta \in \mathrm{bd}([a_i])$, contradicting $\beta \in V \setminus \mathrm{cl}([a_i])$. So $\gamma \in [a_k]$ for some $k < i$. But then $\alpha \in \mathrm{an}(\gamma) = a_k$, which again contradicts $\alpha \in [a_i]$. $\square$

3. The Markov property for reciprocal graphs. From now on, the vertices of the reciprocal graph $G = (V, E)$ will index a set of real-valued random variables, X_α, $\alpha \in V$, defined on some common probability space. So the random vector $X_V = \{X_\alpha : \alpha \in V\}$ assumes values in $\mathbb{R}^V$. The simultaneous probability distribution of X_V is denoted by P, and, for a subset of vertices a, P_{X_a} (or P_a for short) denotes the marginal probability distribution of $X_a = \{X_\alpha : \alpha \in a\}$. Notice that $P_V = P$. For three disjoint subsets a, b and c of V, we will denote the fact that X_a and X_b are conditionally independent given X_c under P by $X_a \perp X_b | X_c$ [P]. Often this will be abbreviated to $a \perp b | c$ [P], or even to $a \perp b | c$ if reference to P is clear. The condition is understood to hold trivially if $a = \varnothing$ or $b = \varnothing$.

For some purposes it is necessary to assume that P satisfies condition (CI5) of Frydenberg (1990a), coined *intersection* by Pearl (1988) and also called the *block independence property* [Whittaker (1990)]:

(BIP)
$$(a \perp b | c \cup d \,[\,P\,] \text{ and } a \perp c | b \cup d \,[\,P\,])$$
$$\text{if and only if } a \perp b \cup c | d \,[\,P\,].$$

A sufficient condition for this is that P is absolutely continuous w.r.t. some product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on the Borel sets of $\mathbb{R}^V$ and has a density which is μ-a.e. positive. (Here and in the sequel each μ_α is assumed σ-finite.) In case the graph is undirected and assuming that the probability measure P fulfills condition (BIP), Pearl and Paz (1986) have proven the equivalence of three types of Markov properties, namely (1) the pairwise Markov property, (2) the local Markov property and (3) the global Markov property. In Frydenberg (1990a) these three Markov properties are generalized to the context of chain graphs and it is shown there that, assuming (BIP), they remain equivalent and can thus be referred to as "the Markov property." It seems only natural to define the three Markov properties for our present still more general context of reciprocal graphs, but it turns out that this does not pay off. Instead, the only concept that we can fruitfully use is that of the global Markov property, so we will turn to its definition now.

The probability measure P is called *global G-Markov* if $a \perp b | c$ [P] whenever c separates a and b in $(G_{\mathrm{an}(a \cup b \cup c)})^m$, where $(G_{\mathrm{an}(a \cup b \cup c)})^m$ is the moral graph of the smallest anterior set containing $a \cup b \cup c$. Notice that this equals Frydenberg's definition verbatim, so if the graph happens to be a chain graph both definitions amount to the same thing. The adjective "global" will often be omitted since this, too, cannot lead to confusion in cases where the graph is a chain graph.

A question that arises is whether the class of probability models determined by reciprocal graphs is strictly larger than the class of probability models determined by chain graphs. Frydenberg (1990a) gives a necessary and sufficient condition for two chain graphs to define the same probability model and it might be the case that for each reciprocal graph G there exists a chain graph $\tilde{G}$ such that P is G-Markov iff P is $\tilde{G}$-Markov. The discussion at the end of Section 5 shows that this is not the case.

It is important to note what the global G-Markov property means in case the graph G is undirected. This is formulated in the next lemma.

LEMMA 3.1. *Suppose $G = (V, E)$ is an undirected graph and P is a probability measure. Then P is global G-Markov if and only if $a \perp b|c \ [P]$ whenever c separates a and b in the graph G.*

PROOF. Since G is undirected, $\operatorname{an}(a \cup b \cup c) = \bigcup\{u(\alpha)|\alpha \in a \cup b \cup c\}$. From this it is easy to see that c separates a and b in G if and only if c separates a and b in $G_{\operatorname{an}(a \cup b \cup c)}$. But $G_{\operatorname{an}(a \cup b \cup c)} = (G_{\operatorname{an}(a \cup b \cup c)})^m$, since G is undirected. $\square$

In Andersson and Perlman (1993) the class of lattice conditional independence (CI) models was introduced in the context of multivariate normal distributions. A lattice CI probability model can be defined as follows. Suppose $A \subseteq \mathscr{P}(V)$ is a finite distributive lattice under the set operations $\cup$ and $\cap$, where $\mathscr{P}(V)$ denotes the power set of V and V indexes a set of real-valued random variables, X_α, $\alpha \in V$, defined on a common probability space. Then the probability measure P satisfies the lattice CI probability model determined by A if, for all $a, b \in A$, $X_{a \setminus b} \perp X_{b \setminus a}|X_{a \cap b} \ [P]$. The condition is understood to be trivially satisfied if $a \subseteq b$ (or $b \subseteq a$). As Andersson and Perlman observe, the lattice CI models differ from graphical (chain) models. There are, however, fundamental relationships with the G-Markov property for reciprocal graphs (see Corollaries 3.5 and 3.6 below). A first impression is given by the following proposition.

PROPOSITION 3.2. *Suppose the probability measure P is G-Markov, where $G = (V, E)$ is a reciprocal graph. Then P satisfies the lattice CI model determined by $A(G)$, the lattice of all anterior sets of G.*

PROOF. Suppose a and b are anterior sets of G such that both $a \setminus b \neq \varnothing$ and $b \setminus a \neq \varnothing$. Since $\operatorname{an}((a \setminus b) \cup (b \setminus a) \cup (a \cap b)) = \operatorname{an}(a \cup b) = a \cup b$, it suffices to show that $a \setminus b$ is separated from $b \setminus a$ by $a \cap b$ in the graph $(G_{a \cup b})^m$; that is, $\{\alpha, \beta\}$ cannot be an edge in $(G_{a \cup b})^m$, whenever $\alpha \in a \setminus b$ and $\beta \in b \setminus a$. So assume $\alpha \in a \setminus b$ and $\beta \in b \setminus a$. Then $\{\alpha, \beta\} \nsubseteq a$ and $\{\alpha, \beta\} \nsubseteq b$. Using Proposition 2.8, it follows that $\{\alpha, \beta\}$ is not an edge of the graph $(G_a)^m \cup (G_b)^m = (G_{a \cup b})^m$. $\square$

LEMMA 3.3. *Suppose the probability measure P is G-Markov, where $G = (V, E)$ is a reciprocal graph. Let $a_1, \ldots, a_q$ be a never-decreasing listing of the elements in $J(A(G))$. Then $[a_i] \perp (\bigcup_{j \leq i} a_j) \setminus \operatorname{cl}([a_i])|\operatorname{bd}([a_i])$ for $1 \leq i \leq q$.*

PROOF. Since $[a_i] \cup (\bigcup_{j \leq i} a_j) \setminus \operatorname{cl}([a_i]) \cup \operatorname{bd}([a_i]) = (\bigcup_{j \leq i} a_j) \ (= b_i,$ say) is an anterior set, it suffices to show that $[a_i]$ is separated from $b_i \setminus \operatorname{cl}([a_i])$ by $\operatorname{bd}([a_i])$ in $(G_{b_i})^m$. But this follows from Proposition 2.9. $\square$

We will now set out for the main result of this section, namely the "Gibbs = Markov" equivalence for reciprocal graphs. Theorem 3.4 below generalizes the corresponding results for chain graphs in Frydenberg (1990a), Theorem 3.3, Corollary 3.4 and Theorem 4.1. The reader is invited to verify that the Gibbs factorization in Theorem 3.4(vii) reduces to the factorization in Frydenberg (1990a), Theorem 4.1(iii), in case the graph is a chain graph. An introduction to the subject of the "Gibbs = Markov" equality can be found in Speed (1979).

REMARK 3.1. Frydenberg's Theorem 4.1 [Frydenberg (1990a)] is there stated and proved under the supposition that the following conjecture holds: "If G is an undirected finite graph, P is a global G-Markov probability measure and P has a strictly positive density p w.r.t. the product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on the Borel sets of $\mathbb{R}^V$, where each μ_α is σ-finite, then p has a Gibbs factorization w.r.t. G." Note that the latter means that $p(x)$ factorizes as $\prod_{c \in K} \psi^c(x_c)$, μ-a.e., where K denotes the set of cliques of G, ψ^c denotes some nonnegative real-valued measurable function on $\mathbb{R}^c$ and $x_c = \{x_\alpha : \alpha \in c\}$. It is not entirely clear if a full proof of the conjecture, known in the literature as the Hammersley–Clifford theorem, has been published somewhere [see, e.g., Isham (1981) and the references therein]. However, as it can be rigorously proven, it is indeed a theorem. [For details, see Koster (1994).] The converse statement, that is, Gibbs factorization implies the global Markov property, is well known [cf. Lauritzen, Dawid, Larsen and Leimer (1990)].

Suppose P, the probability distribution of $\{X_\alpha : \alpha \in V\}$, has a positive density p w.r.t. the product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on the Borel sets of $\mathbb{R}^V$. Let, for $a \subseteq V$, p_a denote (a version of) the marginal density of X_a w.r.t. $\mu_a := \times_{\alpha \in a} \mu_\alpha$ on the Borel sets of $\mathbb{R}^a$. Furthermore, let $p_{a|b}$ denote (a version of) the conditional density of X_a given X_b (i.e., $p_{a|b} = p_{a \cup b}/p_b$, $\mu_{a \cup b}$-a.e.). It is well known that $X_a \perp X_b | X_c$ $[P]$ if and only if $p_{a|b \cup c} = p_{a|c}$, $\mu_{a \cup b \cup c}$-a.e. [cf. Dawid (1979)]. Evaluation of p_a at $x_a := \{x_\alpha : \alpha \in a\}$ and of $p_{a|b}$ at $x_{a \cup b}$ is denoted by $p_a(x_a)$ and $p_{a|b}(x_{a \cup b})$, respectively. The latter notation stresses the fact that $p_{a|b}$ is considered a function of $x_{a \cup b}$. We will sometimes say that $p_{a|b}$ has a Gibbs factorization w.r.t. the undirected graph G. By this we mean that $p_{a|b}(x_{a \cup b})$ factorizes as $\prod_{c \in K} \psi^c(x_c)$, μ-a.e., hence also as $\prod_{c \in K} \psi^c(x_{c \cap (a \cup b)})$, $\mu_{a \cup b}$-a.e. (K denotes the set of cliques of G). In order to prevent cumbersome formulations, we will henceforth mostly omit the "μ-a.e." clauses.

THEOREM 3.4. *If $G = (V, E)$ is a reciprocal graph, P is a probability measure and P has a positive density p w.r.t. the product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on the Borel sets of $\mathbb{R}^V$, then the following statements are equivalent*:

   (i) *P is global G-Markov.*
   (ii) *For each $a \in A(G)$, P_a is global G_a-Markov.*

(iii) *For each $a \in A(G)$, P_a is $(G_a)^m$-Markov.*

(iv) *For each $a \in A(G)$, p_a has a Gibbs factorization w.r.t. $(G_a)^m$.*

(v) *For every never-decreasing listing $a_1, \ldots, a_q$ of the members of $J(A(G))$, $P_{V \setminus [a_q]}$ is $G_{V \setminus [a_q]}$-Markov, P_{a_q} is G_{a_q}-Markov and*

$$[a_q] \perp V \setminus \text{cl}([a_q]) \mid \text{bd}([a_q]) \; [P].$$

(vi) *The density p can be factorized*:

$$p(x) = \prod_{a \in J(A(G))} \prod_{c \in K_a} \psi_a^c(x_{c \cap \text{cl}([a])}), \quad \mu\text{-a.e.}$$

such that $\prod_{c \in K_a} \psi_a^c(x_{c \cap \text{cl}([a])}) = p_{[a] \mid \langle a \rangle}(x_a)$, μ_a-a.e., for each $a \in J(A(G))$.

(vii) *The density p can be factorized*:

$$p(x) = \prod_{a \in J(A(G))} \prod_{c \in K_a} \psi_a^c(x_{c \cap \text{cl}([a])}), \quad \mu\text{-a.e.}$$

such that

$$\int_{\mathbb{R}^{[a]}} \prod_{c \in K_a} \psi_a^c(x_{c \cap \text{cl}([a])}) \, d\mu_{[a]}(x_{[a]}) = 1, \quad \mu_{\langle a \rangle}\text{-a.e.},$$

for each $a \in J(A(G))$.

In (vi) *and* (vii), *K_a denotes the set of cliques in the undirected graph $(G_a)^m$.*

PROOF. We show that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i); (iii) $\Leftrightarrow$ (iv); (ii) $\Rightarrow$ (v) $\Rightarrow$ (iv); (v) $\Rightarrow$ (vi) $\Rightarrow$ (vii) $\Rightarrow$ (iv).

To see that (i) implies (ii), let b, c and d be subsets of the anterior set a and assume that d separates b and c in $(G_{\text{an}(b \cup c \cup d)})^m$, where an$(b \cup c \cup d)$ is taken w.r.t. G_a. From Lemma 2.2 and Proposition 2.3, it follows that it is immaterial whether the set an$(b \cup c \cup d)$ is taken w.r.t. G_a or w.r.t. G. So d separates b and c in $(G_{\text{an}(b \cup c \cup d)})^m$ as well, where now an$(b \cup c \cup d)$ is taken w.r.t. the graph G. Hence $b \perp c \mid d \; [P]$.

Assume (ii). Lemma 3.1 shows that, in order to prove (iii), it suffices to show that $b \perp c \mid d \; [P]$ whenever d separates b and c in the graph $(G_a)^m$. But if d separates b and c in the graph $(G_a)^m$, then [since an$(b \cup c \cup d) \subseteq a$] d separates b and c in the graph $(G_{\text{an}(b \cup c \cup d)})^m$ as well. Hence (ii) implies $b \perp c \mid d \; [P]$.

Assume (iii). To show (i), suppose c separates a and b in $(G_{\text{an}(a \cup b \cup c)})^m$. Since $P_{\text{an}(a \cup b \cup c)}$ is $(G_{\text{an}(a \cup b \cup c)})^m$-Markov, it follows immediately that $a \perp b \mid c \; [P]$.

To show that (iii) and (iv) are equivalent, we refer to Remark 3.1 which establishes that, for each $a \in A(G)$ separately, P_a is $(G_a)^m$-Markov iff p_a has a Gibbs factorization w.r.t. $(G_a)^m$.

Now suppose (ii) holds. The first two statements of (v) are immediate [note that Proposition 2.4(vii) implies that $V \setminus [a_q] \in A(G)$]. The third statement of (v) follows from Lemma 3.3. So (ii) implies (v).

To show that (v) implies (iv), we use induction on $q = |J(A(G))|$. If $q = 1$, then $J(A(G)) = \{V\}$, $[a_q] = V$, so by hypothesis $P = P_V$ is G-Markov which is (i), equivalent to (iv). Assume now the implication holds for all reciprocal graphs H with $|J(A(H))| < q$. Take $a \in A(G)$ fixed. Let $a_1, \ldots, a_q$ be a

never-decreasing listing of the members of $J(A(G))$ such that, if $[a_i] \subseteq a$ and $[a_j] \not\subseteq a$, then $i < j$ [cf. Proposition 2.4(viii)]. If $a \subseteq V \setminus [a_q]$, the first statement of (v) and the already established equivalence of (i) and (iv) (applied to $G_{V \setminus [a_q]}$) show that p_a has a Gibbs factorization w.r.t. $(G_a)^m$. If $a \cap [a_q] \neq \varnothing$, then $[a_q] \subseteq a$, so $[a_i] \subseteq a$ for all $i \leq q$, so $a = V$. We must now show that p has a Gibbs factorization w.r.t. G^m. By hypothesis, $[a_q] \perp V \setminus \mathrm{cl}([a_q])|\mathrm{bd}([a_q])$, hence $[a_q] \perp \langle a_q \rangle \setminus \mathrm{cl}([a_q])|\mathrm{bd}([a_q])$. Then, denoting the conditional density of X_a given X_b by $p_{a|b}$, we obtain for the density p: $p = p_{V \setminus [a_q]} \times p_{[a_q]|V \setminus [a_q]} = p_{V \setminus [a_q]} \times p_{[a_q]|\mathrm{bd}([a_q])} = p_{V \setminus [a_q]} \times p_{[a_q]|\langle a_q \rangle} = p_{V \setminus [a_q]} \times p_{a_q}/p_{\langle a_q \rangle}$. By the induction hypothesis, $p_{V \setminus [a_q]}$ has a Gibbs factorization w.r.t. $(G_{V \setminus [a_q]})^m$, hence also w.r.t. G^m as the latter graph has more edges and hence more and greater cliques. By hypothesis, P_{a_q} is G_{a_q}-Markov, so the already established equivalence of (i) and (iv) shows that p_{a_q} and $p_{\langle a_q \rangle}$ have Gibbs factorizations w.r.t. $(G_{a_q})^m$ and $(G_{\langle a_q \rangle})^m$, respectively, hence also w.r.t. G^m as again the latter graph has more edges and hence more and greater cliques. This shows that p has a Gibbs factorization w.r.t. G^m.

Note that we have now established that (i) to (v) are equivalent.

We proceed by showing that (v) implies (vi). We use induction on $q = |J(A(G))|$. If $q = 1$, then $J(A(G)) = \{V\}$, $a_q = V$. So the hypothesis implies that P is G-Markov; that is, (i) holds. But (i) is equivalent to (iv), so p has a Gibbs factorization w.r.t. G^m. However, since $J(A(G)) = \{V\}, [a_q] = \mathrm{cl}([a_q]) = V$, the factorization in (vi) reduces to $\prod_{c \in K} \psi^c(x_c)$, where K denotes the set of cliques of the undirected graph G^m. This is an "ordinary" Gibbs factorization w.r.t. G^m, which, as has just been shown, holds. Now assume that the implication holds for all reciprocal graphs H with $|J(A(H))| < q$. Since $[a_q] \perp V \setminus \mathrm{cl}([a_q])|\mathrm{bd}([a_q])$, $p = p_{V \setminus [a_q]} \times p_{[a_q]|V \setminus [a_q]} = p_{V \setminus [a_q]} \times p_{[a_q]|\mathrm{bd}([a_q])}$. Since also $[a_q] \perp a_q \setminus \mathrm{cl}([a_q])|\mathrm{bd}([a_q])$, $p_{[a_q]|\mathrm{bd}([a_q])} = p_{[a_q]|\langle a_q \rangle}$, so $p = p_{V \setminus [a_q]} \times p_{[a_q]|\langle a_q \rangle}$. By (v) and induction, $p_{V \setminus [a_q]}(x_{V \setminus [a_q]})$ factorizes as $\prod_{i < q} \prod_{c \in K_{a_i}} \psi_{a_i}^c(x_{c \cap \mathrm{cl}([a_i])})$, where, for $i < q$, $\prod_{c \in K_{a_i}} \psi_{a_i}^c(x_{c \cap \mathrm{cl}([a_i])}) = p_{[a_i]|\langle a_i \rangle}(x_{a_i})$. Furthermore, since P_{a_q} is G_{a_q}-Markov, p_{a_q} has a Gibbs factorization w.r.t. $(G_{a_q})^m$. As (v) is equivalent to (iv), it also follows that $p_{\langle a_q \rangle}$ has a Gibbs factorization w.r.t. $(G_{\langle a_q \rangle})^m$, hence also w.r.t. $(G_{a_q})^m$. So the following equalities hold: $p_{[a_q]|\mathrm{bd}([a_q])}(x_{\mathrm{cl}([a_q])}) = p_{[a_q]|\langle a_q \rangle}(x_{a_q}) = p_{a_q}(x_{a_q})/p_{\langle a_q \rangle}(x_{\langle a_q \rangle}) = \prod_{c \in K_{a_q}} \psi_{a_q}^c(x_c) = \prod_{c \in K_{a_q}} \psi_{a_q}^c(x_{c \cap \mathrm{cl}([a_q])})$, where the last transition is valid since the left-hand side only depends on $x_{\mathrm{cl}([a_q])}$. All in all we have now shown that $p(x) = \prod_{i \leq q} \prod_{c \in K_{a_i}} \psi_{a_i}^c(x_{c \cap \mathrm{cl}([a_i])})$, where, for $i \leq q$, $\prod_{c \in K_{a_i}} \psi_{a_i}^c(x_{c \cap \mathrm{cl}([a_i])}) = p_{[a_i]|\langle a_i \rangle}(x_{a_i})$; that is, (vi) holds.

Clearly, (vi) implies (vii).

Finally, we show that (vii) implies (iv). We again use induction on $q = |J(A(G))|$. If $q = 1$, then $J(A(G)) = \{V\}$, $A(G) = \{\varnothing, V\}$, $[a_q] = \mathrm{cl}([a_q]) = V$, so the factorization of p obtained by the hypothesis is all that is needed to establish (iv). Assume now the implication holds for all reciprocal graphs H with $|J(A(H))| < q$. Take $b \in A(G)$ fixed. We must show that p_b has a Gibbs factorization w.r.t. $(G_b)^m$. Let $a_1, \ldots, a_q$ be a never-decreasing listing of the members of $J(A(G))$ such that, if $[a_i] \subseteq b$ and $[a_j] \not\subseteq b$, then $i < j$ [cf.

Proposition 2.4(viii)]. Suppose $b \subseteq V \setminus [a_q]$. Then

$$
p_{V \setminus [a_q]}(x_{V \setminus [a_q]}) = \int_{\mathbb{R}^{[a_q]}} \prod_{i \leq q} \prod_{c \in K_{a_i}} \psi_{a_i}^c(x_{c \cap \text{cl}([a_i])}) \, d\mu_{[a_q]}(x_{[a_q]})
$$

$$
= \prod_{i < q} \prod_{c \in K_{a_i}} \psi_{a_i}^c(x_{c \cap \text{cl}([a_i])}),
$$

so (vii) is satisfied for $P_{V \setminus [a_q]}$ and $G_{V \setminus [a_q]}$. By the induction hypothesis, (iv) holds for $G_{V \setminus [a_q]}$. So p_b has a Gibbs factorization w.r.t. $(G_b)^m$. If $b \cap [a_q] \neq \varnothing$, then $[a_q] \subseteq b$, so $[a_i] \subseteq b$ for all $i \leq q$, so $b = V$. Clearly, for each $a \in J(A(G))$, $x_a \mapsto \prod_{c \in K_a} \psi_a^c(x_{c \cap \text{cl}([a])})$ is a function with a Gibbs factorization w.r.t. $(G_a)^m$, so also w.r.t. G^m. But then $p(x) = \prod_{a \in J(A(G))} \prod_{c \in K_a} \psi_a^c(x_{c \cap \text{cl}([a])})$ satisfies a Gibbs factorization w.r.t. G^m as well. $\square$

REMARK 3.2.   Note that in proving (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i), we have not used the assumption that P has a positive density w.r.t. μ. In fact, (i), (ii) and (iii) of Theorem 3.4 are equivalent for any probability measure P, as long as it is well understood that, in (iii), "$(G_a)^m$-Markov" means "*global* $(G_a)^m$-Markov."

Theorem 3.4(vi) is very instructive since it offers detailed insight into the meaning of the G-Markov property and its relation to the lattice CI property. Loosely stated, one can say that the Gibbs factorization of p consists of an outer factorization which is determined by the lattice structure of $A(G)$ and an inner factorization which is determined by Gibbs factorizations w.r.t. $(G_a)^m$ for join-irreducible anterior sets a. This is made more precise in the next two corollaries.

COROLLARY 3.5.   *Suppose the probability measure P has a positive density p w.r.t. the product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on the Borel sets of $\mathbb{R}^V$. Let $A \subseteq \mathscr{P}(V)$ be a distributive lattice. Then the following statements are equivalent:*

   (i) *P satisfies the lattice CI probability model determined by A.*
   (ii) *The density p can be factorized: $p(x) = \prod_{a \in J(A)} p_{[a] | \langle a \rangle}(x_a)$, μ-a.e.*

PROOF.   Let $q = |J(A)|$ and let $a_1, \ldots, a_q$ be a never-decreasing listing of the members of $J(A)$. Define the graph $G = (V, E)$ by

$$
E = \bigcup_{a \in J(A)} \mathscr{P}_2([a]) \cup \left(\bigcup_{i < j \leq q} \{(\alpha, \beta) | \alpha \in [a_i], \beta \in [a_j], [a_i] \subseteq a_j\} \right).
$$

Clearly, G is a reciprocal graph (even a chain graph) and $A(G) = A$. Also, for $a \in J(A(G))$, $\text{bd}([a]) = \langle a \rangle$, $\text{cl}([a]) = a$, $(G_a)^m = (a, \mathscr{P}_2(a))$, so the set of cliques in the undirected graph $(G_a)^m$ is $K_a = \{a\}$.

To see that (ii) implies (i), notice that the density p satisfies Theorem 3.4(vi) w.r.t. the graph G [just put $\psi_a^a(x_a) = p_{[a]|\langle a \rangle}(x_a)$, $a \in J(A)$], so P is G-Markov. Proposition 3.2 now implies that P satisfies the lattice CI model determined by $A(G) = A$.

We prove that (i) implies (ii) by induction on $q = |J(A)|$. If $q = 1$, then $A = \{\varnothing, V\}$, $J(A) = \{V\}$, $[V] = V$ and $\langle V \rangle = \varnothing$, so (ii) is trivial. Suppose the implication holds for all lattices with less then q join-irreducible elements. Let $a_1, \ldots, a_q$ be a never-decreasing listing of $J(A)$. From Proposition 2.4(vii) it follows that $V \setminus [a_q] \in A$. Clearly, $P_{V \setminus [a_q]}$ satisfies the lattice CI model determined by $A_{V \setminus [a_q]}$. Notice that $J(A_{V \setminus [a_q]}) = \{a_1, \ldots, a_{q-1}\}$. By induction, $p_{V \setminus [a_q]}(x_{V \setminus [a_q]}) = \prod_{i \le q-1} p_{[a_i]|\langle a_i \rangle}(x_{a_i})$. The lattice CI property for A implies that $a_q \setminus (V \setminus [a_q]) \perp (V \setminus [a_q]) \setminus a_q | a_q \cap (V \setminus [a_q])$, that is, $[a_q] \perp V \setminus a_q | \langle a_q \rangle$. It follows that $p(x) = p_{V \setminus [a_q]}(x_{V \setminus [a_q]}) \times p_{[a_q]|V \setminus [a_q]}(x) = p_{V \setminus [a_q]}(x_{V \setminus [a_q]}) \times p_{[a_q]|\langle a_q \rangle}(x_{a_q}) = \prod_{i \le q} p_{[a_i]|\langle a_i \rangle}(x_{a_i})$. $\square$

In view of Theorem 3.4(vi) and Corollary 3.5 the next result, specifying two further conditions equivalent to the G-Markov condition, should not come as a surprise.

COROLLARY 3.6. *Suppose the probability measure P has a positive density p w.r.t. the product measure $\mu = \times_{\alpha \in V} \mu_\alpha$ on the Borel sets of $\mathbb{R}^V$. Let $G = (V, E)$ be a reciprocal graph. Then the following statements are equivalent:*

(i) *P is G-Markov.*

(ii) *P satisfies the lattice CI probability model determined by $A(G)$, and P_a is G_a-Markov for each $a \in J(A(G))$.*

(iii) *P satisfies the lattice CI probability model determined by $A(G)$; for each $a \in J(A(G))$, $[a] \perp a \setminus \mathrm{cl}([a]) | \mathrm{bd}([a])$ and the conditional density $p_{[a]|\langle a \rangle}$ has a Gibbs factorization w.r.t. $(G_a)^m$.*

PROOF. (i) implies (ii) is clear. Assume (ii) holds. To see (iii), let $a \in J(A(G))$. Note the $\mathrm{bd}([a]) \subseteq a$, so the two final statements of (iii) pertain to the marginal distribution P_a. As P_a is G_a-Markov, both statements follow easily from Theorem 3.4. Finally, assume (iii) holds. By the previous corollary we have the factorization $p(x) = \prod_{a \in J(A(G))} p_{[a]|\langle a \rangle}(x_a)$. Furthermore, for each $a \in J(A(G))$, $p_{[a]|\langle a \rangle}(x_a) = p_{[a]|\mathrm{bd}([a])}(x_{\mathrm{cl}([a])})$. Now $p_{[a]|\langle a \rangle}(x_a)$ has a Gibbs factorization w.r.t. $(G_a)^m$, say $p_{[a]|\langle a \rangle}(x_a) = \prod_{c \in K_a} \psi_a^c(x_c) = \prod_{c \in K_a} (\psi_a^c(x_{c \cap \mathrm{cl}([a])}))$, where the last equality holds since there is only dependence on $x_{\mathrm{cl}([a])}$. Together we obtain the factorization $p(x) = \prod_{a \in J(A(G))} \prod_{c \in K_a} \psi_a^c(x_{c \cap \mathrm{cl}([a])})$, so Theorem 3.4(vi) holds. This means that P is G-Markov. $\square$

Together, Corollaries 3.5 and 3.6 imply that the class of lattice CI probability models with positive densities is a subclass of the class of probability models determined by a reciprocal graph (or even a chain graph). It is easy to see that this inclusion is strict. Recently, without assuming the existence of

positive densities, Andersson, Madigan, Perlman and Triggs (1995a, b) showed that the class of lattice CI probability models coincides with the class of G-Markov probability models, where G is a *transitive* directed acyclic graph (TDAG). Andersson, Madigan, Perlman and Triggs (1995a) also contains Corollary 3.5 and various other conditions equivalent to the lattice CI property.

4. Reciprocal graphs and LISREL models. In this section we will consider multivariate normal probability distributions which obey a set of linear equations. Before doing so we formulate an important lemma concerning general (nondegenerate) normal distributions. It is well known that in this case "independence" is equivalent to "zero correlation." Also, "conditional independence" just means "zero partial correlation." Since partial correlations are closely connected to the inverse of the covariance matrix, the following lemma can be proved [cf. Speed and Kiiveri (1986) and Whittaker (1990)]. Notice that the first three statements of the lemma refer to the simultaneous distribution of all variables, whereas the last three statements pertain to certain marginal distributions.

LEMMA 4.1. *Suppose the random vector $\{X_\alpha : \alpha \in V\}$ has a nondegenerate normal distribution with covariance matrix Σ. Let a, b and c be disjoint subsets of V (a and b nonempty) such that $a \cup b \cup c = V$. Let $\Sigma_{a,b}$ denote the submatrix $(\mathrm{Cov}(X_\alpha, X_\beta))_{\alpha \in a, \beta \in b}$. Then the following statements are equivalent:*

  (i) $X_a \perp X_b | X_c$.
 (ii) $\Sigma_{a,b} = \Sigma_{a,c}(\Sigma_{c,c})^{-1}\Sigma_{c,b}$ $(:= 0$ if $c = \varnothing)$.
(iii) $(\Sigma^{-1})_{a,b} = 0$.
 (iv) *For all $\alpha \in a$ and $\beta \in b$, $X_\alpha \perp X_\beta | X_c$.*
  (v) *For all $\alpha \in a$ and $\beta \in b$, $\mathrm{Cov}(X_\alpha, X_\beta) = \Sigma_{\alpha,c}(\Sigma_{c,c})^{-1}\Sigma_{c,\beta}$ $(:= 0$ if $c = \varnothing)$.*
 (vi) *For all $\alpha \in a$ and $\beta \in b$, $(\mathrm{Cov}(X_\alpha, X_\beta, X_c)^{-1})_{\alpha,\beta} = 0$.*

Let us now turn to linear equations systems. In Jöreskog and Sörbom (1989) the so-called *full LISREL model* is defined by the following equations:

$$\eta = B\eta + \Gamma\xi + \zeta \qquad \text{structural equation}$$

$$\left.\begin{array}{l} Y = \Lambda_y\eta + \varepsilon \\ X = \Lambda_x\xi + \delta \end{array}\right\} \qquad \text{measurement equations}$$

with the assumptions

LI-1. ζ, ε, δ and ξ are mutually uncorrelated;
LI-2. $I - B$ is nonsingular.

A few remarks concerning these assumptions are in order here. First, in LI-1 we make the assumption that ε and ξ are uncorrelated, whereas Jöreskog and Sörbom [(1989), page 4], assume instead that ε and η are uncorrelated. Since $\eta = (I - B)^{-1}(\Gamma\xi + \zeta)$ it is obvious that the latter is

implied by our set of assumptions, but not the other way around. However, since all models in Jöreskog and Sörbom (1989) in fact do have ε uncorrelated with ξ, it seems that this is tacitly presumed by them, so LI-1 simply makes this explicit. Substituting $\eta = (I - B)^{-1}(\Gamma\xi + \zeta)$ into the measurement equation for Y, this also leads to the "natural" condition that in all equations the errors are uncorrelated with the regressors.

The second remark pertains to the assumption that the error variables ζ, ε and δ are mutually uncorrelated. This keeps open the possibility that there are nonzero correlations within the sets of ζ-, ε- and δ-variables. Formulating the full LISREL model as the so-called "Submodel 3B" [cf. Jöreskog and Sörbom (1989), page 157], it even becomes possible to allow correlations between the ε- and δ-variables. This is of course a nice illustration of the flexibility of the LISREL model. It can be argued though that correlated errors are, from a modeling point of view, inelegant, since they introduce correlations between dependent variables which apparently cannot be explained otherwise. In this view, the sole role of error variables is to make both ends of a linear equation meet. They should not be allowed to introduce new relations between the variables of primal interest. Be this as it may, it turns out that the main result in Section 5 can only be proved if all error variables are uncorrelated (in fact, if not, the result is false). Therefore we will make the explicit assumption (i.e., SE-5 below) that all error variables are uncorrelated. However, this assumption is much less restrictive than might be suspected since the case of correlated errors can be handled satisfactorily as the discussion in Section 6 shows.

A final remark concerns the covariance matrix of the exogenous (ξ-) variables, $\Phi = \text{Cov}(\xi)$. In ordinary regression and path analysis models, no special structure is imposed on this covariance matrix: it is only assumed to be strictly positive definite. In LISREL, however, entries of Φ like the entries of all other parameter matrices can be fixed to certain values or constrained to be equal to other parameters. We will see later that too much of this "modeling" of Φ may render some of our results false, so the theorems in Section 5 will only be proved under the supposition (SE-6 below) that Φ is a nondegenerate block-diagonal covariance matrix. Again, the general case with possible violation of this condition is treated in Section 6.

Consider the following way of writing the LISREL model:

$$\begin{bmatrix} Y \\ X \\ \eta \end{bmatrix} = \begin{bmatrix} 0 & 0 & \Lambda_y \\ 0 & 0 & 0 \\ 0 & 0 & B \end{bmatrix} \begin{bmatrix} Y \\ X \\ \eta \end{bmatrix} + \begin{bmatrix} 0 \\ \Lambda_x \\ \Gamma \end{bmatrix} \xi + \begin{bmatrix} \varepsilon \\ \delta \\ \zeta \end{bmatrix}.$$

In this way the model takes the form of a system of $p + q + m$ simultaneous equations. In fact, it has the form of a LISREL "Submodel 2," except for the fact that some of the endogenous and all exogenous variables are latent.

Let us now consider a general (recursive or nonrecursive) simultaneous equations system of p endogenous and q exogenous variables:

$$Y = BY + CX + E,$$

where we assume:

SE-1. $\text{Cov}(E_i, X_j) = 0$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$;
SE-2. $I - B$ is nonsingular;
SE-3. $(Y, X) \sim N_{p+q}(0, \Sigma)$.

Some or all of the Y- and X-variables may be latent, hence the previous formulation of the LISREL model shows that this model is a special case of the present system of equations. Typically, the matrices B and C will be stipulated to have fixed (structural) zeros on certain entries and the same will be the case for various off-diagonal elements of $\Phi := \text{Cov}(X)$ and $\Psi := \text{Cov}(E)$. Let Z_b, Z_c, Z_ϕ and Z_ψ denote the sets of indices of all structural zeros of B, C, Φ and Ψ, respectively. It is assumed that $(i, i) \in Z_b$, $(i, i) \notin Z_\psi$ and $(j, j) \notin Z_\phi$ for $1 \leq i \leq p$ and $1 \leq j \leq q$. Now the matrices B, C, Φ and Ψ are constrained to satisfy the zero structures specified in Z_b, Z_c, Z_ϕ and Z_ψ, respectively; that is, we assume:

SE-4. $Z_b \subseteq Z(B)$, $Z_c \subseteq Z(C)$, $Z_\phi \subseteq Z(\Phi)$ and $Z_\psi \subseteq Z(\Psi)$.

Here $Z(A)$ denotes the set of indices of zero entries of the matrix A.

It is common practice to draw a path diagram of the SE system. As Jöreskog and Sörbom (1989) remark, this has at least two advantages. First, "the path diagram effectively communicates the basic conceptual ideas of the model." Second, "if the path diagram includes sufficient detail, it can represent exactly the corresponding algebraic equations of the model and the [correlational] assumptions about the error terms in these equations." We may add to this that the path diagram also portrays the correlational assumptions about the exogenous variables. But is it possible to extricate more information from it, in particular concerning the conditional independence structure of the random variables?

The following conventions for drawing path diagrams are usually observed:

1. The Y-, X- and E-variables are denoted either by name or short keyword or number; they constitute the vertices of the path diagram.
2. Assumed correlations within the set of X-variables or within the set of E-variables are denoted by two-way arrows. More precisely stated, a two-way arrow between, say, X_i and X_j, $i \neq j$, is drawn if and only if $(i, j) \notin Z_\phi$. Similarly, a two-way arrow between, say, E_i and E_j, $i \neq j$, is drawn if and only if $(i, j) \notin Z_\psi$.
3. One-way arrows always point toward Y-variables. An arrow from X_j to Y_i is drawn if and only if $(i, j) \notin Z_c$; an arrow from Y_j to Y_i is drawn if and only if $(i, j) \notin Z_b$. Finally, one-way arrows point from E_i to Y_i for $i = 1, \ldots, p$.

So we see that path diagrams of SE systems look very similar to reciprocal graphs. Indeed, if the two-way arrows are understood to represent undirected edges, it is clear that there are no one-way arrows between vertices belonging to the same path component, hence path diagrams are really pictures of reciprocal graphs. Now, the question is whether or not the Markov properties

of the random variables involved in the SE system can be read off the path diagram.

There are various reasons why this might not be the case. For one, in the way a path diagram is generated (as outlined above) no reference at all is made to the notion of conditional independence. A path diagram merely portrays a system of linear equations: each variable that has a one-way arrow pointing toward it corresponds to precisely one equation. No such connection with equations, either linear or nonlinear, is present in the case of G-Markov probability measures. Furthermore, the absence of a two-way arrow between two exogenous or error variables means that these variables are uncorrelated. This is just a property of their marginal, bivariate distribution and does not imply anything about them being partially uncorrelated—the linear counterpart of conditional independence (cf. Lemma 4.1).

A simple example may illustrate the fact that the Markov properties (i.e., the conditional independence statements) which can be obtained from a path diagram by interpreting it as a reciprocal graph need not be valid.

EXAMPLE 2. Consider the single equation $Y = c_1 X_1 + c_2 X_2 + c_3 X_3 + E$, together with the assumptions SE-1 to SE-4, where it is supposed that $\mathrm{Cov}(X_1, X_3) = 0$, hence $Z_\phi = \{(1, 3), (3, 1)\}$. Let $\mathscr{P}$ denote the set of normal probability distributions for (Y, X_1, X_2, X_3) satisfying these assumptions. We have the following path diagram:

$$
\begin{array}{l}
X_1 \\
\updownarrow \quad \searrow \\
X_2 \rightarrow Y \leftarrow E \\
\updownarrow \quad \nearrow \\
X_3
\end{array}
$$
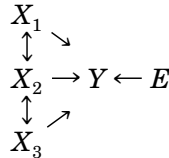
Now, although X_1 and X_3 are separated by X_2 in $(G_{\mathrm{an}(\{X_1, X_2, X_3\})})^m = (\{X_1, X_2, X_3\}, \{\{X_1, X_2\}, \{X_2, X_3\}\})$, it is easy to see that $X_1 \perp X_3 | X_2$ does not hold for all $P \in \mathscr{P}$. On the other hand, although $X_1 \perp X_3$ holds by supposition for all $P \in \mathscr{P}$, it is not true that X_1 and X_3 are separated (by the empty set) in $(G_{\mathrm{an}(\{X_1, X_3\})})^m = (G_{\mathrm{an}(\{X_1, X_2, X_3\})})^m$.

Let us consider the matter in more detail. Since the model in the given example clearly is recursive, at least part of the problem has nothing to do with (non)recursiveness. In order to eliminate inconsistencies as the above, we must introduce two further qualifications [these were also made by Kiiveri, Speed and Carlin (1984), when studying recursive systems of equations]. First, all error variables are assumed to be uncorrelated; second, it is assumed that the exogenous variables can be so ordered that their covariance matrix becomes block-diagonal without structural zero entries within the diagonal blocks [it is not required to actually carry out the reordering, though; the condition simply means that the structural zero entries of $\mathrm{Cov}(X)$ and $\mathrm{Cov}(X)^{-1}$ coincide]. In terms of the sets Z_ψ and Z_ϕ, which contain the indices of structural zero (off-diagonal) entries of $\mathrm{Cov}(E)$ and

$\mathrm{Cov}(X)$, this can be stated as follows:

SE-5. $Z_\psi = E^*(\{1, \ldots, p\})$;

SE-6. If $(i, j) \notin Z_\phi$ and $(j, k) \notin Z_\phi$, then $(i, k) \notin Z_\phi$.

Notice that SE-6 is violated in the previous example since $(1, 3) \in Z_\phi$, whereas $(1, 2) \notin Z_\phi$ and $(2, 3) \notin Z_\phi$. The remarkable fact is that SE-5 and SE-6 suffice to solve all problems; that is, once SE-5 and SE-6 are satisfied, Markov properties can validly be read off a path diagram. This will be shown in the next section.

5. Markov properties of a simultaneous equations system. From now on we will consider a simultaneous equations system with p endogenous and q exogenous variables, $Y = BY + CX + E$, which satisfies assumptions SE-1 to SE-6. The set of normal probability distributions for the random vector (Y, X) satisfying all of this is denoted by $\mathscr{P}$.

We associate with the SE system a graph, G; it is the subgraph obtained from the path diagram by deleting all error variables. More precisely stated, $G = (V, F)$, where $V = \{1, \ldots, p + q\}$ is the set of vertices representing the random variables $Y_1, \ldots, Y_p, X_1, \ldots, X_q$ in this order, while the edge set F consists of the following elements:

1. $(\alpha, \beta) \in F$ iff $1 \le \beta \le p$ and one of the following conditions holds:
   a. $1 \le \alpha \le p$, $\alpha \ne \beta$ and $(\beta, \alpha) \notin Z_b$;
   b. $p + 1 \le \alpha \le p + q$ and $(\beta, \alpha - p) \notin Z_c$.
2. $\{\alpha, \beta\} \in F$ iff $p + 1 \le \alpha, \beta \le p + q$, $\alpha \ne \beta$ and $(\alpha - p, \beta - p) \notin Z_\phi$.

Notice that we leave out all error variables. [We will not discuss extensively the reasons for this exclusion. It suffices to say that they mainly stem from the fact that the vector (Y, X, E) has a degenerate normal distribution: the linear dimension of the $2p + q$ random variables is only $p + q$. This leads to "trivial" conditional independencies such as $Y_i \perp Y_j | \text{rest}$, where "rest" denotes $\{\text{remaining } Y\text{-variables}\} \cup X \cup E$.]

Clearly, the class of undirected path components of G, $U(G)$, is equal to $\{\{1\}, \ldots, \{p\}\} \cup \{u(\alpha) | p + 1 \le \alpha \le p + q\}$. From this it is obvious that there are no directed edges (arrows) between vertices which belong to the same path component. So G is indeed a reciprocal graph; loosely stated, it is just the subgraph obtained from the ordinary path diagram by wiping out all error variables together with their associated arrows.

The next lemma gives well-known expressions for Σ and Σ^{-1} [e.g., see Kiiveri, Speed and Carlin (1984)]. Recall that Σ is the covariance matrix of the random vector (Y, X), $\Phi = \mathrm{Cov}(X)$ and $\Psi = \mathrm{Cov}(E)$.

LEMMA 5.1. *Let $Y = BY + CX + E$ be a simultaneous equations system which satisfies* SE-1 *to* SE-3. *Then*

$$\Sigma = \begin{bmatrix} (I - B)^{-1}(\Psi + C\Phi C')(I - B')^{-1} & (I - B)^{-1}C\Phi \\ \Phi C'(I - B')^{-1} & \Phi \end{bmatrix}$$

and

$$\Sigma^{-1} = \begin{bmatrix} (I - B')\Psi^{-1}(I - B) & -(I - B')\Psi^{-1}C \\ -C'\Psi^{-1}(I - B) & C'\Psi^{-1}C + \Phi^{-1} \end{bmatrix}.$$

We now formulate our main result. Assuming $Z_\phi = E^*(\{1, \ldots, q\})$ in SE-6, that is, uncorrelated exogenous variables, it was discovered independently by Spirtes (1995).

THEOREM 5.2. *Let $Y = BY + CX + E$ be a simultaneous equations system* (*with p endogenous and q exogenous variables*) *that satisfies* SE-1 *to* SE-6. *The set of normal probability distributions for (Y, X) satisfying this system is denoted by $\mathscr{P}$. Let $G = (V, F)$ be the associated reciprocal graph, defined as above. Then all $P \in \mathscr{P}$ are global G-Markov.*

We must show that for disjoint sets a, b and c (a and b nonempty) the implication

c separates a and b in $(G_{\mathrm{an}(a \cup b \cup c)})^m \Rightarrow a \perp b | c \, [P]$ for all $P \in \mathscr{P}$

holds. The proof will be split up in six steps and it will become apparent that in all steps except the last one we are able to show that the implication holds the other way around, too. In fact, we claim that the class $\mathscr{P}$ is *Markov perfect* w.r.t. G [cf. Frydenberg (1990b)], that is,

$a \perp b | c \, [P]$ for all $P \in \mathscr{P} \Rightarrow c$ separates a and b in $(G_{\mathrm{an}(a \cup b \cup c)})^m$

holds, but we are unable to prove it at this point. (However, recent results by Spirtes, Richardson, Meek, Scheines and Glymour (1996) entail Markov perfectness of $\mathscr{P}$. See Section 6 for further discussion of this and related matters.)

PROOF OF THEOREM 5.2. We will first prove the theorem and its converse under the additional supposition that $a \cup b \cup c = V$. Notice that, in this case, c separates a and b in $(G_{\mathrm{an}(a \cup b \cup c)})^m = G^m$ if and only if, for all $\alpha \in a$ and all $\beta \in b$, $\{\alpha, \beta\}$ is not an edge in G^m.

1. Assume that $a = \{\alpha\}$, $b = \{\beta\}$, $c = V \setminus \{\alpha, \beta\}$ and $\{\alpha, \beta\} \subseteq \{1, \ldots, p\}$, that is, both α and β refer to endogenous variables. Then $\alpha \perp \beta | V \setminus \{\alpha, \beta\} \, [P]$ for all $P \in \mathscr{P} \Leftrightarrow$ the (α, β)-entry of Σ^{-1} vanishes for all $P \in \mathscr{P} \Leftrightarrow ((I - B')\Psi^{-1}(I - B))_{\alpha, \beta} = 0$ for all matrices B and Ψ which satisfy the simultaneous equations model [i.e., $Z_b \subseteq Z(B)$ and $Z_\psi \subseteq Z(\Psi)$] $\Leftrightarrow -\psi_{\alpha\alpha}^{-1} b_{\alpha\beta} - \psi_{\beta\beta}^{-1} b_{\beta\alpha} + \sum_{\gamma=1}^p b_{\gamma\alpha} \psi_{\gamma\gamma}^{-1} b_{\gamma\beta} = 0$ for all matrices B and Ψ satisfying $Z_b \subseteq Z(B)$ and $Z_\psi \subseteq Z(\Psi) \Leftrightarrow (\alpha, \beta) \in Z_b$ and $(\beta, \alpha) \in Z_b$ and for all $\gamma \in \{1, \ldots, p\}$ either $(\gamma, \alpha) \in Z_b$ or $(\gamma, \beta) \in Z_b \Leftrightarrow (\beta, \alpha) \notin F$ and $(\alpha, \beta) \notin F$ and there exists no $\gamma \in \{1, \ldots, p\}$ such that both $(\alpha, \gamma) \in F$ and $(\beta, \gamma) \in F \Leftrightarrow \{\alpha, \beta\}$ is not an edge in $G^m \Leftrightarrow V \setminus \{\alpha, \beta\}$ separates $\{\alpha\}$ and $\{\beta\}$ in G^m.

2. Next assume that $a = \{\alpha\}$, $b = \{\beta\}$ and $c = V \setminus \{\alpha, \beta\}$, but now $\alpha \in \{1, \ldots, p\}$ and $\beta \in \{p + 1, \ldots, p + q\}$, so α refers to an endogenous variable whereas β denotes an exogenous variable. Then $\alpha \perp \beta | V \setminus \{\alpha, \beta\} \, [P]$ for all $P \in \mathscr{P} \Leftrightarrow$ the (α, β)-entry of Σ^{-1} is 0 for all $P \in \mathscr{P} \Leftrightarrow -((I - B')\Psi^{-1}C)_{\alpha, \beta - p} = 0$ for all matrices B, C and Ψ which satisfy the simultaneous equations model $\Leftrightarrow -\psi_{\alpha\alpha}^{-1} c_{\alpha, \beta - p} + \Sigma_{\gamma = 1}^{p} \, b_{\gamma\alpha} \psi_{\gamma\gamma}^{-1} c_{\gamma, \beta - p} = 0$ for all matrices B, C and Ψ which satisfy $Z_b \subseteq Z(B)$, $Z_c \subseteq Z(C)$ and $Z_\psi \subseteq Z(\Psi) \Leftrightarrow (\beta, \alpha) \notin F$ and there exists no $\gamma \in \{1, \ldots, p\}$ such that both $(\alpha, \gamma) \in F$ and $(\beta, \gamma) \in F \Leftrightarrow \{\alpha, \beta\}$ is not an edge in $G^m \Leftrightarrow V \setminus \{\alpha, \beta\}$ separates $\{\alpha\}$ and $\{\beta\}$ in G^m.

3. If $a = \{\alpha\}$, $b = \{\beta\}$ and $c = V \setminus \{\alpha, \beta\}$ and both α and β refer to exogenous variables, then $\alpha \perp \beta | V \setminus \{\alpha, \beta\} \, [P]$ for all $P \in \mathscr{P} \Leftrightarrow (C'\Psi^{-1}C + \Phi^{-1})_{\alpha - p, \beta - p} = 0$ for all matrices C, Ψ and Φ which satisfy the simultaneous equations model $\Leftrightarrow (\Phi^{-1})_{\alpha - p, \beta - p} + \Sigma_{\gamma = 1}^{p} c_{\gamma, \alpha - p} \psi_{\gamma\gamma}^{-1} c_{\gamma, \beta - p} = 0$ for all matrices C, Ψ and Φ which satisfy $Z_c \subseteq Z(C)$, $Z_\psi \subseteq Z(\Psi)$ and $Z_\phi \subseteq Z(\Phi) \Leftrightarrow \{\alpha, \beta\} \notin F$ and there exists no $\gamma \in \{1, \ldots, p\}$ such that both $(\alpha, \gamma) \in F$ and $(\beta, \gamma) \in F \Leftrightarrow \{\alpha, \beta\}$ is not an edge in $G^m \Leftrightarrow V \setminus \{\alpha, \beta\}$ separates $\{\alpha\}$ and $\{\beta\}$ in G^m.

4. We proceed by induction on $|a \cup b|$. Suppose $|a \cup b| \geq 3$. Without loss of generality, we may assume that $|b| \geq 2$. Pick some $\beta \in b$. Then, using the (BIP) and induction, $a \perp b | c \, [P]$ for all $P \in \mathscr{P} \Leftrightarrow (a \perp \beta | c \cup b \setminus \{\beta\} \, [P]$ and $a \perp b \setminus \{\beta\} | c \cup \{\beta\} \, [P]$) for all $P \in \mathscr{P}) \Leftrightarrow c \cup b \setminus \{\beta\}$ separates a and $\{\beta\}$ in G^m and $c \cup \{\beta\}$ separates a and $b \setminus \{\beta\}$ in $G^m \Leftrightarrow c$ separates a and b in G^m.

5. Now consider the case that $a \cup b \cup c$ is an arbitrary anterior set, not necessarily V. Suppose $a \cup b \cup c = \mathrm{an}(a \cup b \cup c) = d$, say. Define the sets e and f by $e = d \cap \{1, \ldots, p\}$, $f = d \cap \{p + 1, \ldots, p + q\}$. It is easy to see that the subvector (Y_e, X_f) satisfies a system of $|e|$ simultaneous equations, together with the set of assumptions SE-1 to SE-6 [which are understood to pertain only to (Y_e, X_f)]. Here we use the fact that d is an anterior set, so the subvector (Y_e, X_f) is "self-contained": ignoring errors, no variables but those in (Y_e, X_f) are involved in the definition of its equations. Note that X_f and $X_{\{p + 1, \ldots, p + q\} \setminus d}$ are mutually uncorrelated since d is an anterior set. Furthermore, the set of normal probability distributions for (Y_e, X_f) satisfying SE-1 to SE-6 is just $\mathscr{P}_d = \{P_d | P \in \mathscr{P}\}$. This is true since any subvector (Y_e, X_f) satisfying assumptions SE-1 to SE-6 can always be extended to a random vector, say $(Y_e, \tilde{Y}_{\{1, \ldots, p\} \setminus d}, X_f, \tilde{X}_{\{p + 1, \ldots, p + q\} \setminus d})$, which satisfies the assumptions of the full SE system and has the distribution of (Y_e, X_f) as its marginal distribution. Using the result of step 4 we now get: $a \perp b | c \, [P]$ for all $P \in \mathscr{P} \Leftrightarrow a \perp b | c \, [P_d]$ for all $P_d \in \mathscr{P}_d \Leftrightarrow c$ separates a and b in $(G_d)^m \Leftrightarrow c$ separates a and b in $(G_{\mathrm{an}(a \cup b \cup c)})^m$.

6. Finally, suppose a and b are separated by c in $(G_{\mathrm{an}(a \cup b \cup c)})^m$. Using an argument similar to the one used in step 5, it is clear that we may assume, without loss of generality, that $\mathrm{an}(a \cup b \cup c) = V$. Define $d = V \setminus (a \cup b \cup c)$. Now define d_1 as the set of all vertices $\delta \in d$ for which there exists

a path in G^m to a vertex $\alpha \in a$, such that the path remains outside c [hence the path is contained in the subgraph $(G^m)_{V \setminus c}$]. Now put $d_2 = d \setminus d_1$, then $a \cup d_1$ is separated from $b \cup d_2$ by c in G^m. For suppose there is, say, a path from $\gamma \in d_2$ to $\delta \in d_1$ in $(G^m)_{V \setminus c}$. The definition of d_1 implies the existence of a path in $(G^m)_{V \setminus c}$ from δ to some $\alpha \in a$. Together these paths constitute a path from γ to α (outside c), contradicting the definition of d_2. By a similar argument, a path in $(G^m)_{V \setminus c}$ from $\beta \in b$ to $\delta \in d_1$ can be extended to a path (outside c) from β to some $\alpha \in a$ and hence cannot occur. Finally, paths in $(G^m)_{V \setminus c}$ from $\gamma \in d_2$ to $\alpha \in a$ or from $\beta \in b$ to $\alpha \in a$ are ruled out by definition or by hypothesis. Since $(a \cup d_1) \cup (b \cup d_2) \cup c = V$, it now follows from the previous step that $(a \cup d_1) \perp (b \cup d_2)|c\,[P]$ for all $P \in \mathscr{P}$, hence also $a \perp b|c\,[P]$ for all $P \in \mathscr{P}$. $\square$

REMARK 5.1. It is interesting to consider the following implication of the claim that $\mathscr{P}$ is Markov perfect w.r.t. G. Suppose $a \perp b|c\,[P]$ for all $P \in \mathscr{P}$. Then $(a \cup d_1) \perp (b \cup d_2)|c\,[P]$ for all $P \in \mathscr{P}$, where d_1 and d_2 are defined as in step 6 of the foregoing proof. So the sets of conditionally independent variables a and b can be extended to $a \cup d_1$ and $b \cup d_2$ such that $a \cup b \cup c \cup d_1 \cup d_2$ constitutes an anterior set of the associated reciprocal graph. To appreciate this, it should be understood that, in case of SE systems satisfying SE-1 to SE-6, the concept of an "anterior set" can easily be defined without reference to the associated reciprocal graph. This can be seen as follows. For $1 \le \alpha \le p$, define $Y_\alpha \downarrow = \{Y_\alpha\} \cup \{Y_\beta | 1 \le \beta \le p, (\alpha, \beta) \notin Z_b\} \cup \{X_\beta | 1 \le \beta \le q, (\alpha, \beta) \notin Z_c\}$. Also, for $1 \le \alpha \le q$, let $X_\alpha \downarrow$ denote $\{X_\beta | 1 \le \beta \le q, (\alpha, \beta) \notin Z_\phi\}$. Suppose $A \subseteq \{X_1, \ldots, X_p, Y_1, \ldots, Y_q\}$ and let $a \subseteq V$ be the corresponding subset of vertices in the associated reciprocal graph. Then a is an anterior set if and only if, for all $Z \in A$, $Z \downarrow \subseteq A$.

The partial proof in steps 1–5 of the claim that $\mathscr{P}$ is Markov perfect comprises a result which is worth being formulated separately.

PROPOSITION 5.3. *Let $Y = BY + CX + E$ be a simultaneous equations system that satisfies* SE-1 *to* SE-6, *and let $\mathscr{P}$ denote the set of normal probability distributions for (Y, X) satisfying this system. Let $G = (V, F)$ be the associated reciprocal graph. Suppose a, b and c are disjoint subsets of V (a and b nonempty) such that $a \cup b \cup c$ is an anterior set and assume that $a \perp b|c\,[P]$ for all $P \in \mathscr{P}$. Then a and b are separated by c in $(G_{a \cup b \cup c})^m$.*

EXAMPLE 3 [cf. Whittaker (1990), pages 302 and 303]. Consider the equations $Y_1 = b_{12}Y_2 + c_{11}X_1 + E_1$ and $Y_2 = b_{21}Y_1 + c_{22}X_2 + E_2$, together with assumptions SE-1 to SE-6, where it is assumed that $\text{Cov}(X_1, X_2) = 0$, hence $Z_\phi = \{(1, 2), (2, 1)\}$. Note that $Z_b = \{(1, 1), (2, 2)\}$ and $Z_c = \{(1, 2), (2, 1)\}$. Let $\mathscr{P}$ denote the set of normal probability distributions for (Y_1, Y_2, X_1, X_2) satisfying these assumptions. The associated reciprocal graph is $G = (\{1, 2, 3, 4\}, \{(1, 2), (2, 1), (3, 1), (4, 2)\})$ (recall that the vertices $1, 2, 3, 4$ rep-

resent Y_1, Y_2, X_1, X_2 in this order), that is,

$$3 \rightarrow 1$$
$$\Updownarrow$$
$$4 \rightarrow 2$$

The anterior sets of G are $\varnothing, \{3\}, \{4\}, \{3, 4\}$ and $\{1, 2, 3, 4\}$. Clearly, $\{3\}$ and $\{4\}$ are separated by the empty set in $(G_{\mathrm{an}(\{3, 4\})})^m = (\{3, 4\}, \varnothing)$, hence X_1 and X_2 are marginally independent. Furthermore, G^m is

$$
\begin{array}{cc}
3 \!-\! 1 \\
\times\, | \\
4 \!-\! 2
\end{array}
$$

So $\{3\}$ and $\{4\}$ are separated by $\{1, 2\}$ in G^m. From Theorem 5.2 we can conclude that $X_1 \perp X_2 | (Y_1, Y_2)$ $[P]$ for all $P \in \mathscr{P}$. Indeed, for $\Sigma^{-1} = \mathrm{Cov}(Y, X)^{-1}$ we have

$$
\Sigma^{-1} = \begin{bmatrix}
\psi_{11}^{-1} + \psi_{22}^{-1} b_{21}^2 & -\psi_{11}^{-1} b_{12} - \psi_{22}^{-1} b_{21} & -\psi_{11}^{-1} c_{11} & \psi_{22}^{-1} b_{21} c_{22} \\
-\psi_{11}^{-1} b_{12} - \psi_{22}^{-1} b_{21} & \psi_{22}^{-1} + \psi_{11}^{-1} b_{12}^2 & \psi_{11}^{-1} b_{12} c_{11} & -\psi_{22}^{-1} c_{22} \\
-\psi_{11}^{-1} c_{11} & \psi_{11}^{-1} b_{12} c_{11} & \phi_{11} + \psi_{11}^{-1} c_{11}^2 & 0 \\
\psi_{22}^{-1} b_{21} c_{22} & -\psi_{22}^{-1} c_{22} & 0 & \phi_{22} + \psi_{22}^{-1} c_{22}^2
\end{bmatrix},
$$

so Σ^{-1} has structural zero entries (only) at the positions $(3, 4)$ and $(4, 3)$. We can also illustrate our claim that $\mathscr{P}$ is Markov perfect w.r.t. G. For instance, it is not true that $\{3\}$ and $\{4\}$ are separated by $\{1\}$ in $(G_{\mathrm{an}(\{1, 3, 4\})})^m = G^m$. Indeed, a simple computation, taking, for example, $\Phi = \Psi = I$ and $b_{12} = b_{21} = c_{11} = c_{22} = 1/2$, gives $\sigma_{34} = 0 \neq (1/3) \times (25/9)^{-1} \times (2/3) = \sigma_{31} \sigma_{11}^{-1} \sigma_{14}$, hence it is not true that $X_1 \perp X_2 | Y_1$ for this particular $P \in \mathscr{P}$. $\square$

There are two other issues that can be settled by means of the previous example (I thank anonymous referees for drawing my attention to these points). First, it is well known that different graphs can have the same Markov properties and hence can give rise to the same probability model. Thus far we have not shown that the class of reciprocal graphs, although strictly larger than the class of chain graphs, defines a class of probability models which is strictly larger than the class of models determined by chain graphs. That this is in fact the case follows from the next proposition.

PROPOSITION 5.4. *No chain graph has the same Markov properties as the reciprocal graph* $G = (\{1, 2, 3, 4\}, \{(1, 2), (2, 1), (3, 1), (4, 2)\})$.

PROOF. Clearly, the only two valid, nontrivial conditional independence statements implied by G are $\{3\} \perp \{4\}$ and $\{3\} \perp \{4\} | \{1, 2\}$. Example 3 shows that there exist probability measures that satisfy these two statements only. As there also exist probability measures that satisfy only either one of these statements (and hence are not G-Markov), it follows that a chain graph, say H, will define the same probability model as G if and only if the only

nontrivial graph separation statements which are valid for H are the following:

(A) $\{3\}$ and $\{4\}$ are separated by $\varnothing$ in $(H_{\mathrm{an}(\{3,\,4\})})^m$;
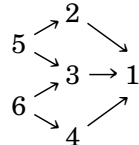(B) $\{3\}$ and $\{4\}$ are separated by $\{1, 2\}$ in H^m.

Suppose $H = (\{1, 2, 3, 4\}, F)$ is a chain graph for which just the statements (A) and (B) hold. It follows immediately that $\{3, 4\} \notin F$, $(3, 4) \notin F$ and $(4, 3) \notin F$. We first show that $\mathrm{an}(\{3, 4\}) = \{3, 4\}$. Suppose $\alpha \in \mathrm{an}(\{3, 4\}) \setminus \{3, 4\} = (\mathrm{an}(3) \cup \mathrm{an}(4)) \setminus \{3, 4\}$, say $\alpha \in \mathrm{an}(3) \setminus \{3, 4\}$. Then there exists a path from 3 to α in $(H_{\mathrm{an}(\{3,\,4\})})^m$. Also, since it is not true that $\{\alpha, 3\}$ and $\{4\}$ are separated by $\varnothing$ in $(H_{\mathrm{an}(\{3,\,4\})})^m$, there exists a path from α to 4 in $(H_{\mathrm{an}(\{3,\,4\})})^m$. Together these paths constitute a path from 3 to 4 in $(H_{\mathrm{an}(\{3,\,4\})})^m$, contradicting statement (A). A similar argument shows that $\alpha \notin \mathrm{an}(4) \setminus \{3, 4\}$. From this it follows that, for all $\alpha \in \{1, 2\}$ and all $\beta \in \{3, 4\}$, $\{\alpha, \beta\} \notin F$ and $(\alpha, \beta) \notin F$. Consequently, $F \subseteq \{(3, 1), (3, 2), (4, 1), (4, 2), e_{12}\}$, where $e_{12} \in \{\{1, 2\}, (1, 2), (2, 1)\}$. This still leaves $2^4 \times 4 = 64$ possibilities for F and hence for H. However, it is impossible that both $(3, 1) \in F$ and $(4, 1) \in F$, as this would contradict statement (B). Similarly, it is impossible that both $(3, 2) \in F$ and $(4, 2) \in F$. It is now an easy exercise to check that, for none of the remaining possibilities for H, (A) and (B) are the only valid graph separation statements. $\square$

Finally, the graph in Example 3 may also serve to illustrate an interesting difference between reciprocal graphs and chain graphs. In chain graphs, every missing edge between two vertices implies a valid, nontrivial conditional independence statement pertaining to these vertices and a certain conditioning set [cf. Wermuth and Lauritzen (1990)]. The entire list of these statements constitutes the content of the so-called *pairwise* Markov property, which is equivalent to the global Markov property for probability measures satisfying (BIP); cf. Frydenberg (1990a), Theorem 3.3. For reciprocal graphs this no longer holds, generally. For instance, although there is no edge between, say, vertices 3 and 2 in Example 3, no valid, nontrivial conditional independence statement concerns these two variables.

6. Markov perfectness and SE systems violating SE-5 or SE-6. We will first briefly discuss SE systems $Y = BY + CX + E$ that satisfy SE-1 to SE-4 but violate SE-5 or SE-6. This issue is treated at length by Spirtes (1995) and Spirtes, Richardson, Meek, Scheines and Glymour (1996) and we refer the reader to these papers for a detailed discussion. In Spirtes (1995) it is shown that the class of all normal probability distributions for SE systems satisfying SE-1 to SE-5, and meeting the additional requirement that $\Phi = \mathrm{Cov}(X)$ is a diagonal matrix, is Markov perfect w.r.t. the associated reciprocal graph (which in this case is a fully directed graph). It turns out that the general case of SE systems satisfying SE-1 to SE-4 but violating SE-5 or SE-6 can essentially be reduced to this situation. Indeed, Spirtes, Richardson, Meek, Scheines and Glymour (1996) show that, by adding a latent variable

for each pair of correlated exogenous or error variables, a so-called "transformed graph" can be made that precisely represents the Markov properties of the original SE system for (Y, X), that is, three subsets of variables satisfy a conditional independence relation (for all probability measures induced by the SE system) if and only if these subsets satisfy a corresponding graph separation condition in the transformed graph. Two examples may help to clarify the procedure.

EXAMPLE 2 (Continued).   Two extra latent variables are added to account for $\mathrm{Cov}(X_1, X_2) \neq 0$ and $\mathrm{Cov}(X_2, X_3) \neq 0$. The transformed graph associated with the SE system for (Y, X), say H, looks as follows. (Notice that the variables Y, X_1, X_2, X_3 are represented by the vertices $1, 2, 3, 4$ respectively; vertices 5 and 6 represent the two added latent variables.)

$$
\begin{array}{c}
\nearrow 2 \searrow \\
5 \searrow \quad \quad \searrow \\
\quad \nearrow 3 \rightarrow 1 \\
6 \searrow \quad \nearrow \\
\quad 4
\end{array}
$$

The anterior sets of H are $\varnothing$, $\{5\}$, $\{6\}$, $\{2, 5\}$, $\{4, 6\}$, $\{5, 6\}$, $\{2, 5, 6\}$, $\{4, 5, 6\}$, $\{3, 5, 6\}$, $\{2, 3, 5, 6\}$, $\{2, 4, 5, 6\}$, $\{3, 4, 5, 6\}$, $\{2, 3, 4, 5, 6\}$, $\{1, 2, 3, 4, 5, 6\}$. Clearly, $\{2\}$ and $\{4\}$ are separated by $\varnothing$ in the graph $(H_{\mathrm{an}(\{2, 4\})})^m = (\{2, 4, 5, 6\}, \{\{2, 5\}, \{4, 6\}\})$. Hence $X_1 \perp X_3$ as it should be. On the other hand it is not true that $\{2\}$ and $\{4\}$ are separated by $\{3\}$ in the graph $(H_{\mathrm{an}(\{2, 3, 4\})})^m = (\{2, 3, 4, 5, 6\}, \{\{2, 5\}, \{3, 5\}, \{3, 6\}, \{4, 6\}, \{5, 6\}\})$, hence $X_1 \perp X_3 \mid X_2$ is not implied by the graph.

EXAMPLE 3 (Continued).   Assume $Z_\psi = \varnothing$, that is, $\mathrm{Cov}(E_1, E_2)$ may be nonzero. We postulate a single latent variable to account for $\mathrm{Cov}(E_1, E_2) \neq 0$. The transformed graph H associated with the SE system is (notice that the variables Y_1, Y_2, X_1, X_2 are represented by the vertices $1, 2, 3, 4$ respectively; vertex 5 represents the added latent variable):

$$
\begin{array}{c}
3 \rightarrow 1 \nwarrow \\
\Updownarrow \quad \searrow 5 \\
4 \rightarrow 2 \swarrow
\end{array}
$$
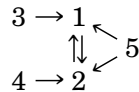
Now, $\mathrm{an}(\{1, 2, 3, 4\}) = \{1, 2, 3, 4, 5\}$ and $H^m = (\{1, \ldots, 5\}, \mathscr{P}_2(\{1, \ldots, 5\}) \setminus \{\{3, 4\}\})$, that is, $\{3, 4\}$ is the only missing edge in H^m. Since $(3, 5, 4)$ is a path in H^m from 3 to 4 which stays outside $\{1, 2\}$, $X_1 \perp X_2 \mid (Y_1, Y_2)$ is *not* implied by the graph. Indeed, using Lemma 5.1 it is easy to see that $(\Sigma^{-1})_{3, 4}$ need not be 0.

As regards Markov perfectness for SE systems satisfying SE-1 to SE-6 we can now state:

THEOREM 6.1.   *Let $Y = BY + CX + E$ be a simultaneous equations system that satisfies* SE-1 *to* SE-6, *and let $\mathscr{P}$ denote the set of normal probability distributions for (Y, X) satisfying this system. Let $G = (V, F)$ be the associated reciprocal graph. Then $\mathscr{P}$ is Markov perfect w.r.t. G.*

PROOF. Let H be the transformed graph associated with the SE system as in Spirtes, Richardson, Meek, Scheines and Glymour (1996). It is easy to see that, under the conditions of the theorem, the transformed graph H and the reciprocal graph G have the same graph separation properties (for subsets of vertices representing (Y, X)). More precisely stated: If a, b and c are disjoint subsets of V, the set of vertices representing $\{Y_1, \ldots, Y_p, X_1, \ldots, X_q\}$, then a and b are separated by c in $(G_{\text{an}(a \cup b \cup c)})^m$ if and only if a and b are separated by c in $(H_{\text{an}(a \cup b \cup c)})^m$. However, the latter condition is equivalent to $a \perp b \mid c \; [P]$ for all $P \in \mathscr{P}$ [cf. Spirtes (1995)]. $\square$

7. Some remaining questions. There are various problems relating to the class of reciprocal graphical models introduced in this paper which remain to be solved. Some of these not only have theoretical interest, but also bear on statistical estimation.

A first question concerns, on the one hand, the set $\mathscr{P}$ of normal probability distributions for (Y, X) satisfying SE-1 to SE-6, and, on the other hand, the set of G-Markov normal probability distributions, say $\mathscr{P}^*$, where G is the reciprocal graph associated with the SE system. According to Theorem 5.2, $\mathscr{P} \subseteq \mathscr{P}^*$. This inclusion may be strict, as can be seen by taking $Z_\phi = \varnothing$ in Example 3, that is, $\text{Cov}(X_1, X_2)$ may be nonzero. The associated reciprocal graph imposes no conditional independence constraints, so any normal probability distribution is in $\mathscr{P}^*$. The LISREL model, on the other hand, still has one degree of freedom.

Second, one would like to have a necessary and sufficient condition to determine when two reciprocal graphs define the same probability model. In Frydenberg (1990a) this question was solved for chain graphs. Two chain graphs have the same Markov properties if and only if they have the same underlying undirected graph and the same "minimal complexes" [Frydenberg (1990a), Theorem 5.6]. The following example shows that this theorem is no longer valid for reciprocal graphs. Let $V = \{1, 2, 3\}$, $G_1 = (V, \{(1, 2), (3, 2)\})$, $G_2 = (V, \{(1, 2), (2, 3), (3, 2)\})$ and $G_3 = (V, \{(1, 2), (2, 3), (3, 1)\})$. Then G_1 and G_2 have the same underlying undirected graph and the same minimal complexes, but have different Markov properties as $1 \perp 3$ is implied by G_1 but not by G_2. On the other hand, G_2 and G_3 have the same Markov properties, but do not have the same underlying undirected graph nor the same minimal complexes. From Corollary 3.6 a sufficient condition can be derived: if two reciprocal graphs, say G and H, have the same join-irreducible anterior sets (and hence also the same anterior sets; cf. Proposition 2.4), and for each join-irreducible anterior set a it holds that $G_a = H_a$, then G and H have the same Markov properties. However, this condition is rather strong and the graphs G_2 and G_3 above show that it is indeed unnecessary.

Recently the problem of Markov equivalence of cyclic directed graphs was addressed (independently) by T. Richardson. Richardson (1994) states a theorem that gives necessary and sufficient conditions for the Markov equivalence of two cyclic directed graphs. Although the conditions can be checked in

polynomial time, it turns out that they are much more complicated than Frydenberg's conditions for chain graphs.

There is a related statistical question, but to address it one first has to think of a proper way to introduce parametric families of distributions within the present context of G-Markov probability models, for example, discrete multivariate cross classifications, multivariate normal distributions, conditional Gaussian (CG) distributions [cf. Lauritzen and Wermuth (1989)] and so on. In view of Theorem 3.4, it seems desirable that, for each join-irreducible anterior set a, both the marginal distribution P_a and the conditional distribution $P_{[a]|\langle a \rangle}$ belong to the parametric family. This means that for a given parametric family only graphs should be considered for which these closure properties hold. Of course, this condition is trivial for discrete multivariate cross classifications and multivariate normal distributions, since these families are closed under marginalization and conditioning anyhow. For other families, for example, CG distributions, the matter may be different. Notice that this requirement of parametric closure under marginalization and conditioning does not entirely agree with the definition of a CG chain model in Wermuth and Lauritzen (1990), since no restrictions on the chain graph are made in that paper. In any case, once this point has been decided upon, the next question is: when do two reciprocal graphs define the same statistical model?

REFERENCES

ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. and TRIGGS, C. M. (1995a). On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *J. Statist. Plann. Inference* **48** 25–46.

ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. and TRIGGS, C. M. (1995b). A graphical characterization of lattice conditional independence models. Unpublished manuscript.

ANDERSSON, S. A. and PERLMAN, M. D. (1993). Lattice models for conditional independence in a multivariate normal distribution. *Ann. Statist.* **21** 1318–1358.

DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539.

DAWID, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31.

DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.

FRYDENBERG, M. (1990a). The chain graph Markov property. *Scand. J. Statist.* **17** 333–353.

FRYDENBERG, M. (1990b). Marginalization and collapsibility in graphical interaction models. *Ann. Statist.* **18** 790–805.

ISHAM, V. (1981). An introduction to spatial point processes and Markov random fields. *Internat. Statist. Rev.* **49** 21–43.

JÖRESKOG, K. G. and SÖRBOM, D. (1989). *LISREL 7. A Guide to the Program and Applications*, 2nd ed. SPSS Inc., Chicago.

KIIVERI, H. and SPEED, T. P. (1982). Structural analysis of multivariate data: a review. In *Sociological Methodology 1982* (S. Leinhardt, ed.). Jossey-Bass, San Francisco.

KIIVERI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30–52.

KOSTER, J. T. A. (1994). Gibbs-factorization and the Markov property. Unpublished manuscript.

LAURITZEN, S. L. (1979). *Lectures on Contingency Tables*. Inst. Mathematical Statistics, Univ. Copenhagen.

LAURITZEN, S. L. (1989). Mixed graphical association models. *Scand. J. Statist.* **16** 273–306.

LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57.

LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H.-G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491–505.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

PEARL, J. and PAZ, A. (1986). Graphoids. A graph-based logic for reasoning about relevancy relations. In *Proceedings of the Seventh European Conference on Artificial Intelligence*. Brighton, UK.

RICHARDSON, T. (1994). Properties of cyclic graphical models. MS thesis, Carnegie Mellon Univ.

SPEED, T. P. (1979). A note on nearest-neighbour Gibbs and Markov probabilities. *Sankhyā Ser. A* **41** 184–197.

SPEED, T. P. and KIIVERI, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14** 138–150.

SPIRTES, P. (1995). Directed cyclic graphical representation of feedback models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (P. Besnard and S. Hanks, eds.). Morgan Kaufmann, San Mateo, CA.

SPIRTES, P., RICHARDSON, T., MEEK, C., SCHEINES, R. and GLYMOUR, C. (1996). Using *d*-separation to calculate zero partial correlations in linear models with correlated errors. Technical Report CMU-Phil-72, Carnegie Mellon Univ.

WERMUTH, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.* **75** 963–972.

WERMUTH, N. and LAURITZEN, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. Roy. Statist. Soc. Ser. B* **52** 21–50.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, New York.

WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Statist.* **5** 161–215.

ERASMUS UNIVERSITY ROTTERDAM
P.O. BOX 1738
3000 DR ROTTERDAM
THE NETHERLANDS
E-MAIL: koster@soc.fsw.eur.nl