NORMAL LINEAR REGRESSION MODELS WITH RECURSIVE GRAPHICAL MARKOV STRUCTURE*

Steen A. Andersson and Michael D. Perlman†

Indiana University and University of Washington

14 June 1998

Abstract

A multivariate normal statistical model defined by the Markov properties determined by an acyclic digraph admits a recursive factorization of its likelihood function (LF) into the product of conditional LFs, each factor having the form of a classical multivariate linear regression model (≡ MANOVA model). Here these models are extended in a natural way to normal linear regression models whose LFs continue to admit such recursive factorizations, from which maximum likelihood estimators and likelihood ratio (LR) test statistics can be derived by classical linear methods. The central distribution of the LR test statistic for testing one such multivariate normal linear regression model against another is derived, and the relation of these regression models to block-recursive normal linear systems is established. It is shown how a collection of nonnested dependent normal linear regression models (≡ seemingly unrelated regressions) can be combined into a single multivariate normal linear regression model by imposing a parsimonious set of graphical Markov (≡ conditional independence) restrictions.

1. Introduction.

Graphical Markov models use graphs, either undirected, directed, or mixed, to represent multivariate statistical dependencies. Statistical variables are represented by the nodes of the graph, then local dependencies are specified by postulating that each variable is conditionally independent of all other variables given its *neighbors* (for undirected graphs), or conditionally independent of its *nondescendants* given its *parents* (for directed graphs). Although the local dependencies may be relatively simple, complex multivariate dependencies are determined by the global structure of the graph. The statistical aspects of graphical Markov models are surveyed in the books by Whittaker (1990), Edwards (1995), Lauritzen (1996), and Cox and Wermuth (1996) and in Cox and Wermuth (1993).

Graphical Markov models given by acyclic digraphs (ADGs), also called recursive Markov models, ADG Markov models, or simply ADG models, have especially amenable statistical properties. The joint probability density function (pdf) of an ADG model factors according to the graph into the product of the conditional pdfs for each variable given its parents, substantially reducing the dimensionality of the parameter space. A clear treatment of the general mathematical properties of ADG models can be found in Lauritzen *et al* (1990). The statistical analysis of ADG models was treated by Wermuth (1980), Kiiveri, Speed and Carlin (1984), and Shachter and Kenley (1989) for continuous

* This research was supported in part by the U. S. National Science Foundation.

† *Address for correspondence*: Department of Statistics, Box 35-4322, University of Washington, Seattle, WA 98195-4322, USA. *e-mail*: michael@stat.washington.edu.

multivariate distributions, by Wermuth and Lauritzen (1983) for discrete distributions, and by Lauritzen and Wermuth (1989) for distributions with both continuous and discrete components.

ADG models allow efficient computational algorithms for exact probability calculations and efficient updating algorithms for Bayesian analysis, hence have been widely used for the construction of expert systems and for causal modelling. These aspects of ADG models can be found in the papers by Lauritzen and Spiegelhalter (1988) and Spiegelhalter *et al* (1993) and in the books by Pearl (1988), Neapolitan (1990), Oliver and Smith (1990), Spirtes *et al* (1993), and Almond (1995).

In this paper we study multivariate statistical models that combine the ADG Markov property with multivariate linear regression, focussing primarily on the normal ($\equiv$ Gaussian) case. Most recursive multivariate normal models studied previously have concentrated only on the covariance structure, with only very simple structure (e.g., MANOVA structure, see Definition 6.1), if any, assumed for the regression ($\equiv$ mean-value) subspace (e.g., Lauritzen and Wermuth (1989, Section 6)). For such simple linear regression models, it is well-known that the joint likelihood function (LF) factors according to the graph into the product of conditional LFs corresponding to lower-dimensional linear regression models (in fact, this is true regardless of the mean-value assumptions), and, furthermore, that the joint parameter space factors into the product of the parameter spaces associated with these lower-dimensional models.

We shall address the following question: under the Markov covariance structure determined by an ADG D, what is the largest class of linear regression ($\equiv$ mean-value) subspaces L for which the joint parameter space continues to factor according to D into the product of the parameter spaces associated with the family of conditional LFs? Our answer, presented in Section 6, is the class of *D-linear* subspaces, whose structure is characterized in Sections 6 and 10 and illustrated by a series of examples in Section 13. Like the classical MANOVA regression models, these normal linear ADG models are amenable in the sense that each conditional LF and associated parameter space has the form of a classical multivariate normal linear regression model ($\equiv$ MANOVA model) and therefore can be solved by standard linear methods to yield explicit maximum likelihood estimators (MLE) and likelihood ratio (LR) test statistics.

For example, by imposing the Markov covariance restrictions determined by a suitable ADG D, it is possible to formulate the following non-standard multivariate linear regression model as a linear ADG model: a 4-variate two-way MANOVA model with no interactions, with no row or column effects for variable 1, no column effects for variable 2, and no row effects for variable 3 (cf. Example 10 in Section 13).

After reviewing basic graph-theoretic terminology in Section 2, in Section 3 we discuss the basic definition and properties of ADG models, including the construction of general ADG submodels via ADG homomorphisms. In Sections 4 and 5 we add the assumption of multivariate normality and review the covariance structure of these normal ADG models. In Section 6 these models are extended by introducing the fundamental class of multivariate D-linear regression subspaces, which are further characterized in Section 10. Maximum likelihood estimators for the resulting class of normal D-linear ADG models are obtained and studied in Sections 7 and 8. The general problem of testing one such multivariate normal linear ADG model against another is treated in Section 9, including the derivation of the LR test statistic and its central distribution. The relation between normal linear ADG models and block-recursive normal linear systems is established in Section 11. In Section 12 we show how a collection of nonnested dependent linear regression models ($\equiv$ Zellner's (1962) seemingly unrelated regression (SUR) model) can be combined into a single parsimonious normal linear ADG model, then extend this to the case where it is desired to test one such SUR model against another. Section 13 contains a series of examples illustrating the preceding ideas.

The class of normal linear ADG models includes the generalized MANOVA models and totally ordered normal linear models ([AMP] (1993)), as well as the normal linear lattice conditional independence (LCI) models introduced in [AP] (1994) (cf. Remarks 4.1, 9.3, 10.1, 10.3, 12.2 and Proposition 11.2 below). The results in this paper may be regarded as extensions of those in [AP] (1993, 1994, 1995a)

2. Acyclic digraphs (ADGs).

A *directed graph (digraph)* D is a pair (V, R), where V is a finite set of *vertices* and $R \subseteq (V \times V) \setminus \Delta$ is a binary relation (the set of *directed edges*) on V such that $(u, v) \in R$ implies $(v, u) \notin R$. Here, $\Delta \equiv \Delta(V)$ is the diagonal $\{(v, v) | v \in V\}$; thus loops and multiple edges are excluded from D. We use the customary arrow $u \rightarrow v$ in our figures to indicate that $(u, v) \in R$, but in the text this relation is indicated by $u \prec_D v$, or simply by $u \prec v$ when D is understood. The corresponding reflexive relation $\bar{R} := R \cup \Delta$ is denoted by $\preceq_D$ (or simply by $\preceq$). Thus $u \preceq v$ means that $u \prec v$ or $u = v$.

We write $u < v \equiv u <_D v$ if $u \prec v$ or there exist $v_1, \ldots, v_k \in V$, $k \geq 1$, such that $u \prec v_1 \prec \cdots \prec v_k \prec v$. The relation $<$ is transitive. The corresponding reflexive relation is denoted by $\leq \, \equiv \, \leq_D$.

An *acyclic digraph* (ADG) is a directed graph $D \equiv (V, R)$ with the property that $v \not< v$ for all $v \in V$. Here the relation $\leq_D$ is a partial ordering on V, i.e., it is reflexive, antisymmetric, and transitive. Every ADG D admits a *never-decreasing listing* (not necessarily unique) of its vertices: $V = \{v_1, \ldots, v_r\}$ where $i < j \Rightarrow v_j \not\leq v_i$.

For an ADG $D \equiv (V, R)$ and $v \in V$, define $\mathrm{pa}(v) := \{u \in V | u \prec v\}$, the *parents* of v; $\mathrm{an}(v) := \{u \in V | u < v\}$, the *ancestors* of v; $\mathrm{de}(v) := \{u \in V | v < u\}$, the *descendants* of v; and $\mathrm{nd}(v) := \{u \in V | v \not\leq u\} \equiv V \setminus (\mathrm{de}(v) \cup \{v\})$, the *nondescendants* of v. Note that $\mathrm{pa}(v) \subseteq \mathrm{nd}(v)$ and that $\mathrm{an}(v)$, $\mathrm{de}(v)$, and $\mathrm{nd}(v)$ depend on the relation $\prec$ only through the corresponding partial ordering $<$. A set $A \subseteq V$ is *ancestral* if $\mathrm{an}(v) \subseteq A$ for all $v \in A$; again, the definition of an ancestral set depends on $\prec$ only through $<$. The set $\mathbf{A}(D)$ of all ancestral subsets of V forms a ring of subsets, the *ancestral ring* of D.

Let $E \equiv (W, S)$ and $D \equiv (V, R)$ be two ADGs. A mapping $f : W \rightarrow V$ is an *ADG homomorphism* if $(f \times f)(S \cup \Delta(W)) \subseteq R \cup \Delta(V)$, i.e., if $w_1 \preceq_E w_2$ implies that $f(w_1) \preceq_D f(w_2)$ for all $w_1, w_2 \in W$. Such an ADG homomorphism is denoted as $f : E \rightarrow D$; see §13.3 for examples. We say that f is a *proper* ADG homomorphism if there exist distinct $w', w'' \in W$ such that $w' \not\preceq_E w''$ and $w'' \not\preceq_E w'$ but $f(w') \preceq_D f(w'')$. If $W = V$ and $S \subseteq R$ then the identity mapping $\mathrm{id}_V : V \rightarrow V$ is an ADG homomorphism from E to D; it is proper if $S \subset R$. If $f : E \rightarrow D$ is an ADG homomorphism, then f is also a poset homomorphism from W to V endowed with the partial orderings $\leq_E$ and $\leq_D$, respectively.

3. Markov models determined by acyclic digraphs (ADGs).

Let $D \equiv (V, R)$ be an ADG. We consider multivariate probability distributions P on a product probability space $\mathbf{X} := \times(\mathbf{X}_v | v \in V)$, where each $\mathbf{X}_v$ is a measurable space sufficiently regular to ensure the existence of regular conditional probabilities. Such a distribution P is conveniently represented by a random variate $x := (x_v | v \in V) \in \mathbf{X}$. For any subset $A \subseteq V$, define $x_A := (x_v | v \in A)$, so $x = x_V$ and $x_\emptyset :=$ constant. We often abbreviate x_v and x_A by v and A, respectively.

For three pairwise disjoint subsets A, B, C of V, we write

$$A \perp\!\!\!\perp B \mid C\,[P] \tag{3.1}$$

to indicate that x_A and x_B are conditionally independent given x_C under P. Trivially, $A \perp\!\!\!\perp B \mid C\,[P]$ if $A = \emptyset$ or $B = \emptyset$, while $A \perp\!\!\!\perp B \mid \emptyset\,[P]$ means that $A \perp\!\!\!\perp B\,[P]$. We require the following elementary property of conditional independence (cf. Dawid (1980)). If A, B, C, F are pairwise disjoint subsets of V, then

$$A \perp\!\!\!\perp B \mid F\,[P] \text{ and } A \perp\!\!\!\perp C \mid F \cup B\,[P] \iff A \perp\!\!\!\perp B \cup C \mid F\,[P]. \tag{3.2}$$

Definition 3.1. A probability measure P on $\mathbf{X}$ is *(local) D-Markovian* if

$$v \perp\!\!\!\perp \big(\mathrm{nd}_D(v) \setminus \mathrm{pa}_D(v)\big) \mid \mathrm{pa}_D(v)\,[P] \;\; \forall\, v \in V. \tag{3.3}$$

Lauritzen *et al* (1990, Proposition 4) define the *global* Markov property determined by D and show that it is equivalent to the local Markov property for ADGs; thus the global property need not be considered separately here. If $\{v_1, \ldots, v_r\}$ is a never-decreasing listing of V, it follows from Proposition 5 of Lauritzen *et al* (1990) that P is D-Markovian iff

$$v_m \perp\!\!\!\perp \big(\{v_1, \ldots, v_{m-1}\} \setminus \mathrm{pa}_D(v_m)\big) \mid \mathrm{pa}_D(v_m)\,[P], \quad m = 2, \ldots, r. \tag{3.4}$$

The *Markov model* $\mathcal{P}(D) \equiv \mathcal{P}(D; \mathbf{X})$ determined by D and $\mathbf{X}$, or, simply, the *ADG model* $\mathcal{P}(D)$, is defined to be the family of all D-Markovian distributions P on $\mathbf{X}$. Note that if $\breve{D} := (V, \breve{R})$ where $\breve{R} \subseteq R$ (i.e., $\breve{D}$ has fewer edges than D), then the acyclic property of D implies that

$$\mathrm{pa}_{\breve{D}}(v) \subseteq \mathrm{pa}_D(v) \subseteq \mathrm{nd}_D(v) \subseteq \mathrm{nd}_{\breve{D}}(v), \tag{3.5}$$

for every $v \in V$, hence $\mathcal{P}(\breve{D}) \subseteq \mathcal{P}(D)$.

In general, sub-ADG models $\mathcal{P}(E; \mathbf{X})$ of $\mathcal{P}(D; \mathbf{X})$ arise in the following way. Let $E \equiv (W, S)$ be a second ADG, $(\mathbf{Y}_w | w \in W)$ a second family of regular measurable spaces, and $\psi\colon E \to D$ a surjective ADG homomorphism such that

$$\mathbf{X}_v = \times(\,\mathbf{Y}_w \mid w \in W, \psi(w) = v) \;\; \forall v \in V, \tag{3.6}$$

so that

$$\mathbf{X} := \times(\,\mathbf{X}_v \mid v \in V) = \times(\,\mathbf{Y}_w \mid w \in W). \tag{3.7}$$

By (3.7), a distribution P on $\mathbf{X}$ can also be represented by a random variate $y := (y_w | w \in W)$. For any subset $B \subseteq W$, define $y_B := (y_w | w \in B)$ and abbreviate y_B by B. By (3.6), $x_A = y_{\psi^{-1}(A)}$ for any $A \subseteq V$.

Proposition 3.1. (i) $\mathcal{P}(E; \mathbf{X})$ is a sub-ADG model of $\mathcal{P}(D; \mathbf{X})$, i.e., $\mathcal{P}(E) \subseteq \mathcal{P}(D)$.
(ii) If also $\psi \colon E \to D$ is a proper ADG homomorphism and each measurable space $\mathbf{Y}_w$, $w \in W$, admits a non-degenerate probability distribution, then $\mathcal{P}(E; \mathbf{X})$ is a proper sub-ADG model of $\mathcal{P}(D; \mathbf{X})$, i.e., $\mathcal{P}(E) \subset \mathcal{P}(D)$.

Proof. (i) Let $v_1, \ldots, v_r$ be a never-decreasing listing of V and, for each $m = 1, \ldots, r$, let $w_{m1}, \ldots, w_{mq_m}$ be a never-decreasing listing of $\psi^{-1}(v_m)$. It follows from the order-preserving property of ψ that the combined sequence $w_{11}, \ldots, w_{1q_1}, \ldots, w_{r1}, \ldots, w_{rq_r}$ is a never-decreasing listing of W. Now define the ADG $\tilde{E} \equiv (W, \tilde{S})$ as follows:

$$\begin{aligned} w_{ml} &\prec_{\tilde{E}} w_{mk} \text{ if } l < k, \\ w_{nl} &\prec_{\tilde{E}} w_{mk} \text{ if } n \neq m \text{ and } v_n \prec_D v_m. \end{aligned}$$

Since E and $\tilde{E}$ differ only in that $S \subseteq \tilde{S}$ (E has fewer edges than $\tilde{E}$), $\mathcal{P}(E) \subseteq \mathcal{P}(\tilde{E})$. We shall complete the proof by showing that $\mathcal{P}(D) = \mathcal{P}(\tilde{E})$.

The following relations are immediate from the construction of $\tilde{E}$. For $m = 1, \ldots, r$ and $k = 1, \ldots, q_m$,

$$\text{pa}_{\tilde{E}}(w_{mk}) = \{w_{m1}, \ldots, w_{m(k-1)}\} \dot{\cup} \psi^{-1}\big(\text{pa}_D(v_m)\big), \tag{3.8}$$

$$\text{nd}_{\tilde{E}}(w_{mk}) = \{w_{m1}, \ldots, w_{m(k-1)}\} \dot{\cup} \psi^{-1}\big(\text{nd}_D(v_m)\big), \tag{3.9}$$

hence

$$\big(\text{nd}_{\tilde{E}}(w_{mk}) \setminus \text{pa}_{\tilde{E}}(w_{mk})\big) = \psi^{-1}\big(\text{nd}_D(v_m) \setminus \text{pa}_D(v_m)\big). \tag{3.10}$$

Suppose first that $P \in \mathcal{P}(D)$. For $m = 1, \ldots, r$ and $k = 1, \ldots, q_m$, apply (3.3) with $v = v_m$ and (3.10) to obtain

$$v_m \perp\!\!\!\perp \big(\text{nd}_{\tilde{E}}(w_{mk}) \setminus \text{pa}_{\tilde{E}}(w_{mk})\big) \mid \text{pa}_D(v_m)\, [P],$$

so

$$\{w_{m1}, \ldots, w_{m(k-1)}, w_{mk}\} \perp\!\!\!\perp \big(\text{nd}_{\tilde{E}}(w_{mk}) \setminus \text{pa}_{\tilde{E}}(w_{mk})\big) \mid \text{pa}_D(v_m)\, [P].$$

It follows that

$$w_{mk} \perp\!\!\!\perp \big(\text{nd}_{\tilde{E}}(w_{mk}) \setminus \text{pa}_{\tilde{E}}(w_{mk})\big) \mid \{w_{m1}, \ldots, w_{m(k-1)}\} \dot{\cup} \text{pa}_D(v_m)\, [P],$$

so that

$$w_{mk} \perp\!\!\!\perp \big(\text{nd}_{\tilde{E}}(w_{mk}) \setminus \text{pa}_{\tilde{E}}(w_{mk})\big) \mid \text{pa}_{\tilde{E}}(w_{mk})\, [P] \tag{3.11}$$

by (3.8), hence $P \in \mathcal{P}(\tilde{E})$.

Conversely, suppose that $P \in \mathcal{P}(\tilde{E})$, so that (3.11) holds for $m = 1, \ldots, r$ and $k = 1, \ldots, q_m$. By (3.10),

$$w_{mk} \perp\!\!\!\perp \big(\mathrm{nd}_D(v_m) \setminus \mathrm{pa}_D(v_m)\big) \mid \mathrm{pa}_{\tilde{E}}(w_{mk})\,[P]. \tag{3.12}$$

The two relations obtained from (3.12) with $k = q_m - 1$ and $k = q_m$ combine according to (3.2) to yield

$$\{w_{m(q_m-1)}, w_{mq_m}\} \perp\!\!\!\perp \big(\mathrm{nd}_D(v_m) \setminus \mathrm{pa}_D(v_m)\big) \mid \mathrm{pa}_{\tilde{E}}(w_{m(q_m-1)})\,[P]. \tag{3.13}$$

Combine this relation with that obtained from (3.12) with $k = q_m - 2$, then continue this process for $k = q_m - 3, \ldots, 1$, finally obtaining

$$\{w_{m1}, \ldots, w_{mq_m}\} \perp\!\!\!\perp \big(\mathrm{nd}_D(v_m) \setminus \mathrm{pa}_D(v_m)\big) \mid \mathrm{pa}_{\tilde{E}}(w_{m1})\,[P]. \tag{3.14}$$

Since $\{w_{m1}, \ldots, w_{mq_m}\} = \psi^{-1}(v_m)$ and $\mathrm{pa}_{\tilde{E}}(w_{m1}) = \mathrm{pa}_D(v_m)$ by (3.8), (3.14) $\equiv$ (3.3), hence $P \in \mathcal{P}(D)$.

(ii) Let $y \equiv (y_w | w \in W) \in \times(\mathbf{Y}_w | w \in W) \equiv \mathbf{X}$ be a random variate such that $(y_w | w \neq w', w'')$ (w', w'' as in Section 2), are mutually independent and independent of $(y_{w'}, y_{w''})$, while $y_{w'}$ and $y_{w''}$ are dependent with non-degenerate distributions; denote the distribution of y by P. By hypothesis, either $w'' \in \mathrm{nd}_E(w') \setminus \mathrm{pa}_E(w')$ or $w' \in \mathrm{nd}_E(w'') \setminus \mathrm{pa}_E(w'')$. Since

$$w' \not\!\perp\!\!\!\perp w'' \mid \mathrm{pa}_E(w')\,[P]$$

and

$$w'' \not\!\perp\!\!\!\perp w' \mid \mathrm{pa}_E(w'')\,[P],$$

it follows from (3.3) for $E \equiv (W, R)$ that $P \notin \mathcal{P}(E)$.

Next, either $\psi(w') = \psi(w'')$ or $\psi(w') \prec_D \psi(w'')$. In the first case, it follows immediately from (3.6) and the definition of P that for each $v \in V$, v, $\mathrm{nd}_D(v) \setminus \mathrm{pa}_D(v)$, and $\mathrm{pa}_D(v)$ are mutually independent, hence P trivially satisfies (3.3), so $P \in \mathcal{P}(D)$. The same argument holds in the second case except for $v = \psi(w'')$, but then $\psi(w') \in \mathrm{pa}_D(v)$, hence

$$(v, \mathrm{pa}_D(v)) \perp\!\!\!\perp \big(\mathrm{nd}_D(v) \setminus \mathrm{pa}_D(v)\big).$$

By (3.2), v satisfies (3.3), so again $P \in \mathcal{P}(D)$. This completes the proof.

Lauritzen *et al* (1990) have characterized the class of D-Markovian distributions P that are absolutely continuous with respect to a product measure $\mu := \otimes(\mu_v | v \in V)$ on $\mathbf{X}$, where each μ_v is a σ-finite measure on $\mathbf{X}_v$. They say that P admits a *D-recursive factorization* if P admits a probability density function of the form

$$p(x) = \prod (k^v(x_v, x_{\mathrm{pa}(v)}) | v \in V), \;\; x \in \mathbf{X},$$

for some measurable functions $k^v \geq 0$ on $\mathbf{X}_v \times \mathbf{X}_{\mathrm{pa}(v)}$ such that $\int k^v(x_v, x_{\mathrm{pa}(v)}) d\mu_v(x_v) = 1$ for all $x_{\mathrm{pa}(v)} \in \mathbf{X}_{\mathrm{pa}(v)}$ and all $v \in V$.

Proposition 3.2. (Lauritzen *et al* (1990, Theorem 1)). Assume that P is absolutely continuous with respect to the product measure μ on $\mathbf{X}$. Then P is D-Markovian if and only if P admits a D-recursive factorization. In this case, $k^v(\cdot, x_{\mathrm{pa}(v)})$ is a version of the conditional density $p(\cdot|x_{\mathrm{pa}(v)})$ of x_v given $x_{\mathrm{pa}(v)}$ and the D-recursive factorization of P assumes the following form:

$$p(x) = \prod \big(p(x_v | x_{\mathrm{pa}(v)}) \bigm| v \in V \big), \; x \in \mathbf{X}. \tag{3.15}$$

Remark 3.1. If $(k^v | v \in V)$ is a family of functions satisfying the above conditions, then

$$\bar{p}(x) := \prod \big(k^v(x_v, x_{\mathrm{pa}(v)}) \bigm| v \in V \big), \; x \in \mathbf{X},$$

defines a probability density wrt μ on $\mathbf{X}$. To verify that $\int \bar{p} d\mu = 1$, choose a never-decreasing listing $v_1, \ldots, v_r$ of V and integrate $\bar{p}$ iteratively over $\mathbf{X}_{v_m}$ wrt μ_{v_m}, $m = r, \ldots, 1$. Thus by Proposition 3.2, $\bar{P} := \bar{p} \cdot \mu$ is D-Markovian.

4. Normal ADG models.

For the remainder of this paper, $D \equiv (V, R)$ shall denote an ADG as in the preceding section, but we now add the assumption that $\mathbf{X}_v = \mathbf{R}^{I_v}$, $v \in V$, where the I_v are nonempty finite index sets, so that $\mathbf{X} = \mathbf{R}^I$ with $I := \dot\cup (I_v | v \in V)$. The *normal ADG model* $\mathbf{N}_I(D)$ is defined[1] to be the restriction of the Markov model $\mathcal{P}(D; \mathbf{R}^I)$ to the class of nonsingular multivariate normal distributions on $\mathbf{R}^I$. In this section we characterize $\mathbf{P}(D; I)$, the set of $I \times I$ positive definite covariance matrices corresponding to the normal ADG model $\mathbf{N}_I(D)$.

For any subset $J \subseteq I$ and vector $x \equiv (x_i | i \in I) \in \mathbf{R}^I$, define $x_J := (x_i | i \in J) \in \mathbf{R}^J$. For any subset $A \subseteq V$, define $I_A := \dot\cup (I_v | v \in A)$ and $x_A := x_{I_A} \in \mathbf{R}^{I_A}$. Note that $x \equiv x_I \equiv x_V$ and define $x_\emptyset := 0$.

For any subsets $J, K \subseteq I$, let $\mathbf{M}(J \times K)$ denote[2] the vector space of all real $J \times K$ matrices, $\mathbf{P}(J)$ the cone of all real positive definite $J \times J$ matrices, and set $\mathbf{M}(J) := \mathbf{M}(J \times J)$. Denote the $J \times J$ identity matrix by 1_J. For $\Sigma \in \mathbf{P}(I)$, let Σ_{JK} denote the $J \times K$ submatrix of Σ, let $\Sigma_J := \Sigma_{JJ} \in \mathbf{P}(J)$, and let Σ_J^{-1} denote $(\Sigma_J)^{-1}$. For disjoint subsets $J, K \subseteq I$, define

$$\Sigma_{J \bullet K} := \Sigma_J - \Sigma_{JK} \Sigma_K^{-1} \Sigma_{KJ} \; \in \mathbf{P}(J).$$

For $v \in V$, define the following three disjoint subsets of I:

$$\not\prec v \not\succ := I_{\mathrm{nd}(v) \setminus \mathrm{pa}(v)}, \qquad \prec v \succ := I_{\mathrm{pa}(v)}, \qquad [v] := I_v;$$

[1] Note that $\mathbf{N}_I(D)$ depends on the partitioning $I = \dot\cup (I_v | v \in V)$ as well as on I itself.

[2] If $J = \emptyset$ or $K = \emptyset$, then $\mathbf{M}(J \times K) := \{0\}$, the zero vector space. If $J \neq \emptyset$ and $K \neq \emptyset$, the product of a $J \times \emptyset$ matrix and a $\emptyset \times K$ matrix is the $J \times K$ zero matrix.

then define[3]

$$\preceq v \succeq \; := \; \prec v \succ \dot{\cup} \; [v].$$

Thus for $\Sigma \in \mathbf{P}(I)$, we have the following partitioning:

$$\Sigma_{\not\prec v \not\succ \dot{\cup} \prec v \succ \dot{\cup} [v]} = \begin{pmatrix} \Sigma_{\not\prec v \not\succ} & \Sigma_{\not\prec v \succ} & \Sigma_{\not\prec v]} \\ \Sigma_{\prec v \not\succ} & \Sigma_{\prec v \succ} & \Sigma_{\prec v]} \\ \Sigma_{[v \not\succ} & \Sigma_{[v \succ} & \Sigma_{[v]} \end{pmatrix}, \tag{4.1}$$

where $\Sigma_{\not\prec v \not\succ} \in \mathbf{P}(\not\prec v \not\succ)$, $\Sigma_{\prec v \succ} \in \mathbf{P}(\prec v \succ)$, $\Sigma_{[v]} \in \mathbf{P}([v])$, $\Sigma_{\prec v \not\succ} := \Sigma_{\prec v \succ \not\prec v \not\succ} \in \mathbf{M}(\prec v \succ \times \not\prec v \not\succ)$, $\Sigma_{[v \not\succ} := \Sigma_{[v] \not\prec v \not\succ} \in \mathbf{M}([v] \times \not\prec v \not\succ)$, $\Sigma_{[v \succ} := \Sigma_{[v] \prec v \succ} \in \mathbf{M}([v] \times \prec v \succ)$, and $\Sigma_{\not\prec v \succ} = (\Sigma_{\prec v \not\succ})^t$, $\Sigma_{\not\prec v]} = (\Sigma_{[v \not\succ})^t$, $\Sigma_{\prec v]} = (\Sigma_{[v \succ})^t$. Furthermore, define

$$\Sigma_{[v]\bullet} := \Sigma_{[v]\bullet \prec v \succ} \in \mathbf{P}([v]). \tag{4.2}$$

and recall that $|\Sigma_{[v]\bullet}| = \frac{|\Sigma_{\preceq v \succeq}|}{|\Sigma_{\prec v \succ}|}$, where $|\Sigma| := \det(\Sigma)$.

Definition 4.1. For $\Sigma \in \mathbf{P}(I)$, the family of matrices

$$\pi_D(\Sigma) := \big((\Sigma_{[v \succ} \Sigma^{-1}_{\prec v \succ}, \Sigma_{[v]\bullet}) \;\big|\; v \in V \big) \;\in\; \times \big(\mathbf{M}([v] \times \prec v \succ) \times \mathbf{P}([v]) \;\big|\; v \in V \big) \;=:\; \Pi(D; I)$$

is called the family of *D-parameters* of Σ.

Let $\mathcal{N}_I(\xi, \Sigma)$ denote the normal distribution on $\mathbf{R}^I$ with mean vector $\xi \in \mathbf{R}^I$ and covariance matrix $\Sigma \in \mathbf{P}(I)$. Then $\mathcal{N}_I(\xi, \Sigma) \in \mathcal{P}(D; \mathbf{R}^I)$ $\forall \xi \in \mathbf{R}^I$ iff $\mathcal{N}_I(0, \Sigma) \in \mathcal{P}(D; \mathbf{R}^I)$.

Definition 4.2. The subset $\mathbf{P}(D; I) \subseteq \mathbf{P}(I)$ is defined as follows:

$$\Sigma \in \mathbf{P}(D; I) \iff \mathcal{N}_I(0, \Sigma) \in \mathcal{P}(D; \mathbf{R}^I);$$

that is, $\Sigma \in \mathbf{P}(D; I)$ iff $x_{[v]} \perp\!\!\!\perp x_{\not\prec v \not\succ} \mid x_{\prec v \succ}$ $\forall v \in V$ when $x \sim \mathcal{N}_I(0, \Sigma)$.

Every $\Sigma \in \mathbf{P}(D; I)$ is uniquely determined by its D-parameters $\pi_D(\Sigma)$:

Proposition 4.1. The mapping

$$\pi_D \colon \mathbf{P}(D; I) \to \Pi(D; I) \tag{4.3}$$

is bijective.

[3] Since $\prec v \succ, \not\prec v \not\succ$, and $\preceq v \succeq$ depend on the ADG $D \equiv (V, R)$ through R as well as V, when necessary we will denote them as $\prec v \succ_D, \not\prec v \not\succ_D$, and $\preceq v \succeq_D$, respectively. The same is true of the indices $[v \succ$, $[v \not\succ$, $\prec v \not\succ$, etc. in (4.1). It is also important to bear in mind the dependence of all these quantities on the partitioning $I = \dot{\cup}(I_v | v \in V)$, and therefore the same dependence of quantities such as $\Sigma_{[v]\bullet} \equiv \Sigma_{[v]\bullet \prec v \succ}$, $\pi_D(\Sigma)$, $\Pi(D; I)$, $\bar{\pi}_D(\xi, \Sigma)$, and $\overline{\Pi}(L, D; I)$ (cf. (4.2) and Definitions 4.1 and 6.3 and Proposition 6.1).

Proof. For any $\big((\beta_v, \Lambda_v) \;\big|\; v \in V\big) \in \Pi(D; I)$, apply Remark 3.1 with $\mu_v :=$ Lebesgue measure on $\mathbf{R}^{[v]}$ and

$$k^v(x_{[v]}, x_{\prec v\succ}) := \big(d\mathcal{N}_{[v]}(\beta_v x_{\prec v\succ}, \Lambda_v)/d\mu_v\big)(x_{[v]}),\ x \in \mathbf{R}^I,$$

to see that

$$\overline{p}(x) := \prod\Big(\big(d\mathcal{N}_{[v]}(\beta_v x_{\prec v\succ}, \Lambda_v)/d\mu_v\big)(x_{[v]}) \;\Big|\; v \in V\Big),\ x \in \mathbf{R}^I \tag{4.4}$$

is a probability density function wrt $\mu :=$ Lebesgue measure on $\mathbf{R}^I$. Since $\overline{p}(x) = c \cdot \exp\big\{-\frac{1}{2}Q(x)\big\}$ for some positive semidefinite quadratic form on $\mathbf{R}^I$, necessarily Q is positive definite, hence $Q(x) = \mathrm{tr}\big(\Sigma^{-1}xx^t\big)$ for some unique $\Sigma \in \mathbf{P}(I)$ and $c^{-1} = (2\pi)^{|I|/2}|\Sigma|^{1/2}$; here $|I| := \mathrm{card}(I)$ and $|\Sigma| := \det(\Sigma)$. Set $x = 0$ in (4.4) to obtain

$$|\Sigma| = \prod\big(\,|\Lambda_v| \;\big|\; v \in V\big), \tag{4.5}$$

which combines with (4.4) to yield

$$\mathrm{tr}\big(\Sigma^{-1}xx^t\big) = \sum\Big(\,\mathrm{tr}\big(\Lambda_v^{-1}(x_{[v]} - \beta_v x_{\prec v\succ})(x_{[v]} - \beta_v x_{\prec v\succ})^t\big) \;\Big|\; v \in V\Big),\ x \in \mathbf{R}^I. \tag{4.6}$$

It follows from Proposition 3.2 that $\mathcal{N}_I(0, \Sigma) \equiv \overline{p} \cdot \mu \in \mathcal{P}(D; \mathbf{R}^I)$, hence $\Sigma \in \mathbf{P}(D; I)$. Furthermore, by (3.15) and the well-known fact that

$$x_{[v]} \mid x_{\prec v\succ} \sim \mathcal{N}_{[v]}\big(\,\Sigma_{[v\succ}\Sigma^{-1}_{\prec v\succ}x_{\prec v\succ}, \Sigma_{[v]\bullet}\big) \tag{4.7}$$

when $x \sim \mathcal{N}_I(0, \Sigma)$, we have $\beta_v = \Sigma_{[v\succ}\Sigma^{-1}_{\prec v\succ}$ and $\Lambda_v = \Sigma_{[v]\bullet}, v \in V$, hence π_D is surjective.

Next suppose that $\Sigma, \Sigma' \in \mathbf{P}(D; I)$ satisfy $\pi_D(\Sigma) = \pi_D(\Sigma')$. By (4.7), $\mathcal{N}_I(0, \Sigma)$ and $\mathcal{N}_I(0, \Sigma')$ determine the same family of conditional distributions, hence $\mathcal{N}_I(0, \Sigma) = \mathcal{N}_I(0, \Sigma')$ by (3.15). Therefore $\Sigma = \Sigma'$, so π_D is injective. This completes the proof.

The covariance set $\mathbf{P}(D; I)$ was defined indirectly according to a probabilistic criterion in Definition 4.2. Relations (ii) and (iii) in the following proposition give two direct algebraic characterizations of $\mathbf{P}(D; I)$. Note too that the inverse mapping π_D^{-1} in (4.3) is implicitly determined by (iii).

Proposition 4.2. The following three conditions are equivalent:

(i) $\Sigma \in \mathbf{P}(D; I)$;

(ii) $\forall v \in V,\ \Sigma_{[v\not\succ} = \Sigma_{[v\succ}\Sigma^{-1}_{\prec v\succ}\Sigma_{\prec v\not\succ}$;

(iii) $\forall x \in \mathbf{R}^I$,

$$\mathrm{tr}\big(\Sigma^{-1}xx^t\big) = \sum\Big(\mathrm{tr}\big(\Sigma^{-1}_{[v]\bullet}(x_{[v]} - \Sigma_{[v\succ}\Sigma^{-1}_{\prec v\succ}x_{\prec v\succ})(x_{[v]} - \Sigma_{[v\succ}\Sigma^{-1}_{\prec v\succ}x_{\prec v\succ})^t\big) \;\Big|\; v \in V\Big).$$

Furthermore, when Σ satisfies these conditions, it also satisfies

$$|\Sigma| = \prod \big(|\Sigma_{[v]\bullet}| \;\big|\; v \in V \big). \tag{4.8}$$

Proof. The equivalence of (i) and (ii) is clear. If Σ satisfies (i), then (iii) and (4.8) follow from (4.6) and (4.5), respectively, by setting $\big((\beta_v, \Lambda_v) \;\big|\; v \in V \big) = \pi_D(\Sigma)$ in the proof of Proposition 4.1. Conversely, if Σ satisfies (iii), then (i) follows from Proposition 3.2.

Remark 4.1. Results similar to Proposition 4.1 in varying degrees of generality have appeared in Kiiveri *et al* (1984, §3), Wermuth (1992, Proposition 4.1), and in unpublished notes of P. S. Eriksen.

By comparing Proposition 4.2 with Theorem 2.1 of [AP] (1993), it is seen that the class of normal lattice conditional independence (LCI) models introduced in [AP] (1993) is a subclass of the class of all normal ADG models. In fact [AMPT] (1995, 1997) have shown (without the assumption of normality) that the class of all LCI models is a proper subclass of all ADG models, namely, those determined by *transitive* ADGs (also see Section 10).

Similarly, it is well-known that every Markov model determined by a *decomposable* undirected graph coincides with some ADG model, but not conversely - see [AMP] (1997a). Therefore, under the assumption of multivariate normality, the class of decomposable *covariance selection models* (Dempster (1972), Lauritzen (1996)) is a proper subclass of the class of normal ADG models.

5. Reconstruction of the covariance matrix from its D-parameters.

It will be shown in Section 7 that for the normal ADG model $\mathbf{N}_I(D)$, the maximum likelihood estimate (MLE) $\hat{\Sigma}$ of the covariance matrix $\Sigma \in \mathbf{P}(D; I)$ is obtained in terms of the MLEs of its D-parameters $\pi_D(\Sigma)$. In order to recover $\hat{\Sigma}$, an explicit representation of the inverse π_D^{-1} of the mapping (4.3) is needed. This representation is now described by the following *Reconstruction Algorithm*, a generalization of the algorithm given in [AP] (1993, Section 2.7).

Let $v_1, \ldots, v_r$ be a never-decreasing listing of the elements in V. For notational convenience abbreviate v_m by m, $[v_m]$ by $[m]$, $\prec v_m \succ$ by $\prec m \succ$, and $[v_m \succ$ by $[m \succ$. Partition Σ according to the decomposition

$$I = [1]\dot{\cup} \cdots \dot{\cup}[r] \tag{5.1}$$

and list the D-parameters $\pi_D(\Sigma)$ in the corresponding order:

$$\pi_D(\Sigma) =: \big((\beta_m, \Lambda_m) \;\big|\; m = 1, \ldots, r \big) \in \times \big(\mathbf{M}([m] \times \prec m \succ) \times \mathbf{P}([m]) \;\big|\; m = 1, \ldots, r \big).$$

(Note that $\mathbf{M}([1] \times \prec 1 \succ) = \{0\}$ and $\beta_1 = 0$.)

Step 1 : $$\Sigma_{[1]} = \Lambda_1.$$

Step 2 : $$\Sigma_{[2\succ} = \beta_2 \Sigma_{\prec 2 \succ}$$ $$\Sigma_{[2]} = \Lambda_2 + \beta_2 \Sigma_{\prec 2]}.$$

Since $\prec\!2\!\succ\ \subseteq\ [1]$, the submatrix $\Sigma_{[1]\dot\cup[2]}$ is now completely determined.

Because $\prec 3 \succ\ \subseteq [1]\dot\cup[2]$, $\Sigma_{\prec 3\succ}$ is a submatrix of $\Sigma_{[1]\dot\cup[2]}$ and the next step may be carried out.

Step $3a$:
$$\begin{aligned}\Sigma_{[3\succ} &= \beta_3\Sigma_{\prec 3\succ}\\ \Sigma_{[3]} &= \Lambda_3 + \beta_3\Sigma_{\prec 3]}.\end{aligned}$$

It is important to note that after Steps 1, 2, and 3a, the two submatrices $\Sigma_{[1]\dot\cup[2]}$ and $\Sigma_{\prec 3\succ\dot\cup[3]}$ are now determined but the complete matrix $\Sigma_{[1]\dot\cup[2]\dot\cup[3]}$ may not yet be fully determined. Since
$$([1]\dot\cup[2])\backslash\!\prec\!3\!\succ\ \subseteq\ \not\prec\!3\!\not\succ,$$
the remaining $[3]\times\big(([1]\dot\cup[2])\backslash\prec 3\succ\big)$ submatrix of $\Sigma_{[1]\dot\cup[2]\dot\cup[3]}$, denoted by $\Sigma_{[3\rangle}$, is determined from $\Sigma_{[1]\dot\cup[2]}$ by means of Proposition 4.2(ii):

Step $3b$:
$$\Sigma_{[3\rangle} = \Sigma_{[3\succ}\Sigma^{-1}_{\prec 3\succ}\Sigma_{\prec 3\rangle} = \beta_3\Sigma_{\prec 3\rangle},$$

where $\Sigma_{\prec 3\rangle}$ is the $\prec\!3\!\succ\times\big(([1]\dot\cup[2])\backslash\!\prec\!3\!\succ\big)$ submatrix of $\Sigma_{[1]\dot\cup[2]}$.

After $m-1$ such steps, the submatrix $\Sigma_{[1]\dot\cup\cdots\dot\cup[m-1]}$ is fully determined and in turn may be used to obtain $\Sigma_{[1]\dot\cup\cdots\dot\cup[m]}$ as follows. First note that the never-decreasing nature of $v_1,\dots,v_r$ implies that

$$\prec\!m\!\succ\ \subseteq\ [1]\dot\cup\cdots\dot\cup[m-1], \tag{5.2}$$
$$([1]\dot\cup\cdots\dot\cup[m-1])\backslash\!\prec\!m\!\succ\ \subseteq \not\prec\!m\!\not\succ\ . \tag{5.3}$$

Thus, if we denote the $[m]\times\big(([1]\dot\cup\cdots\dot\cup[m-1])\backslash\!\prec\!m\!\succ\big)$ submatrix of $\Sigma_{[1]\dot\cup\cdots\dot\cup[m]}$ by $\Sigma_{[m\rangle}$ and the $\prec m\succ\ \times\big(([1]\dot\cup\cdots\dot\cup[m-1])\backslash\prec m\succ\big)$ submatrix by $\Sigma_{\prec m\rangle}$, it follows from (5.2) that both $\Sigma_{\prec m\succ}$ and $\Sigma_{\prec m\rangle}$ are submatrices of $\Sigma_{[1]\dot\cup\cdots\dot\cup[m-1]}$, so that the next step may be carried out to fully determine $\Sigma_{[1]\dot\cup\cdots\dot\cup[m]}$:

Step m :
$$\begin{aligned}\Sigma_{[m\succ} &= \beta_m\Sigma_{\prec m\succ}\\ \Sigma_{[m]} &= \Lambda_m + \beta_m\Sigma_{\prec m]}\\ \Sigma_{[m\rangle} &= \Sigma_{[m\succ}\Sigma^{-1}_{\prec m\succ}\Sigma_{\prec m\rangle} = \beta_m\Sigma_{\prec m\rangle}.\end{aligned}$$

The last equation follows from (5.3) and Proposition 4.2(ii).

The Reconstruction Algorithm is complete after r steps.

6. Normal linear ADG models.

In this section, the classical multivariate normal linear regression model ($\equiv$ MANOVA model) is extended to the *normal linear ADG model* that incorporates the Markov covariance structure determined by an acyclic digraph (ADG) $D \equiv (V,R)$. Such a generalized multivariate linear regression model retains many of the properties of the classical

MANOVA model. In particular, its likelihood function (LF) factors into a product of conditional LFs, each corresponding to a MANOVA model, so that likelihood inference can be carried out by the usual linear methods. The notation and terminology of Section 4 is continued here.

First we briefly review the classical MANOVA model $\mathbf{N}_{I\times N}(L)$, where I and N are finite index sets with $n := |N|$ and where $L \subseteq \mathbf{M}(I\times N)$ is a MANOVA subspace. These results also appear in [AMP] (1993).

Definition 6.1. A linear subspace $L \subseteq \mathbf{M}(I \times N)$ is called a *MANOVA subspace* if $\mathbf{M}(I)L \subseteq L$ or, equivalently (since $1_I \in \mathbf{M}(I)$), if $\mathbf{M}(I)L = L$.

It can be shown that L is a MANOVA subspace of $\mathbf{M}(I\times N)$ iff

$$L = \mathbf{M}(I\times N)P \tag{6.1}$$

for some (necessarily unique) $N \times N$ orthogonal projection matrix $P \equiv P_L$ ($P^t = P$, $P^2 = P$). For fixed I, this establishes a 1-1 corespondence between all MANOVA subspaces $L \subseteq \mathbf{M}(I\times N)$ and all $N\times N$ projection matrices P. Note that $\dim(L) = |I| \cdot \mathrm{tr}(P_L)$.

In tensor product notation, $\mathbf{M}(I\times N) = \mathbf{R}^I \otimes \mathbf{R}^N$, and L is a MANOVA subspace of $\mathbf{M}(I\times N)$ iff

$$L = \mathbf{R}^I \otimes K \tag{6.2}$$

for some (necessarily unique) linear subspace $K \equiv K_L \subseteq \mathbf{R}^N$. For fixed I, this establishes a 1-1 corespondence between all MANOVA subspaces $L \subseteq \mathbf{M}(I\times N)$ and all linear subspaces $K \subseteq \mathbf{R}^N$. It follows from (6.1) and (6.2) that $K_L = \mathbf{R}^N P_L$, or equivalently, that K_L is the row space of P_L:

$$K_L = \mathrm{row}(P_L). \tag{6.3}$$

Remark 6.1. Any linear subspace $L \subseteq \mathbf{M}(I\times N)$ of the form

$$L = \{\, BZ \mid B \in \mathbf{M}(I\times T)\,\} \tag{6.4}$$

is a MANOVA subspace, where T is a finite index set and $Z \in \mathbf{M}(T\times N)$ is a design matrix. Here, $K_L = \mathrm{row}(Z)$ and $P_L = Z^t(ZZ^t)^- Z$, where $(ZZ^t)^-$ denotes any generalized inverse of ZZ^t. Conversely, every MANOVA subspace can be represented (non-uniquely) in the form (6.4).

Each MANOVA subspace $L \subseteq \mathbf{M}(I \times N)$ uniquely determines a normal MANOVA model $\mathbf{N}_{I\times N}(L)$ defined as follows:

$$\mathbf{N}_{I\times N}(L) := \big(\mathcal{N}_{I\times N}(\xi, \Sigma \otimes 1_N) \,\big|\, (\xi, \Sigma) \in L\times \mathbf{P}(I)\big). \tag{6.5}$$

This statistical model consists of n independent normal random vectors

$$x_j \sim \mathcal{N}_I(\xi_j, \Sigma),\ j \in N,$$

with $\xi_j \in \mathbf{R}^I$ and $\Sigma \in \mathbf{P}(I)$, where $\xi := (\xi_j | j \in N) \in L$. If

$$y := (x_j | j \in N) \in \mathbf{M}(I \times N)$$

denotes an observation from this model, then it is well known (cf. Anderson (1984, Chapter 8)) that the maximum likelihood estimator (MLE) $(\hat{\xi}(y), \hat{\Sigma}(y)) \in L \times \mathbf{P}(I)$ is unique and exists for a.e. y [Lebesgue] if and only if

$$n \geq \mathrm{tr}(P_L) + |I| \equiv \dim(K_L) + |I|. \tag{6.6}$$

In this case the MLEs are given by

$$\hat{\xi}(y) = yP_L, \quad n\hat{\Sigma}(y) = yQ_L y^t, \tag{6.7}$$

where $Q_L := 1_N - P_L$, and the maximum of the likelihood function is

$$|\hat{\Sigma}(y)|^{-n/2} \exp\{-n|I|/2\}. \tag{6.8}$$

The MLE $(\hat{\xi}, \hat{\Sigma})$ is a complete and sufficient statistic for the MANOVA model $\mathbf{N}_{I \times N}(L)$.

In the classical MANOVA model $\mathbf{N}_{I \times N}(L)$, the covariance matrix $\Sigma \in \mathbf{P}(I)$ is unrestricted. If, instead, the assumption that $\Sigma \in \mathbf{P}(D; I)$ is imposed, then the class of MANOVA subspaces can be replaced by the larger class of *D-linear subspaces* L (see Definition 6.2). We shall see that the resulting *normal linear ADG model*

$$\mathbf{N}_{I \times N}(L, D) := \big(\mathcal{N}_{I \times N}(\xi, \Sigma \otimes 1_N) \,\big|\, (\xi, \Sigma) \in L \times \mathbf{P}(D; I)\big) \tag{6.9}$$

retains most of the amenable features of the classical model. See §13.1 and §13.3 for examples of such linear ADG models.

For any $\xi \in \mathbf{M}(I \times N)$ and any subset $J \subseteq I$, let ξ_J denote the $J \times N$ submatrix of ξ. For any linear subspace $L \subseteq \mathbf{M}(I \times N)$ and any $v \in V$, define the two subspaces

$$\begin{aligned} L_{[v]} &:= \{\xi_{[v]} | \xi \in L\} \quad \subseteq \mathbf{M}([v] \times N), \\ L_{\prec v \succ} &:= \{\xi_{\prec v \succ} | \xi \in L\} \subseteq \mathbf{M}(\prec v \succ \times N). \end{aligned}$$

Thus we may consider the natural embedding

$$\begin{aligned} L &\to \times(L_{[v]} | v \in V) \subseteq \mathbf{M}(I \times N) \\ \xi &\mapsto (\xi_{[v]} | v \in V). \end{aligned} \tag{6.10}$$

Definition 6.2. A linear subspace $L \subseteq \mathbf{M}(I \times N)$ is called a *D-linear subspace*, or simply a *D-subspace*, if it satisfies the following three conditions:

(i) $L = \times(L_{[v]} | v \in V)$, i.e., (6.10) is a bijection;

(ii) $\forall v \in V$, $L_{[v]}$ is a MANOVA subspace of $\mathbf{M}([v] \times N)$;

(iii) $\forall v \in V, \mathbf{M}([v] \times \prec v \succ) L_{\prec v \succ} \subseteq L_{[v]}$.

Remark 6.2. By (6.1) and (6.2), condition (iii) may be restated in the following two equivalent forms, whose significance is discussed in Section 10:

(iii)′ $\forall u, v \in V$ with $u \prec v$, $P_{L_{[v]}} P_{L_{[u]}} = P_{L_{[u]}}$.

(iii)″ $\forall u, v \in V$ with $u \prec v$, $K_{L_{[u]}} \subseteq K_{L_{[v]}}$.

Definition 6.3. For $(\xi, \Sigma) \in \mathbf{M}(I \times N) \times \mathbf{P}(I)$, the family of matrices

$$\begin{aligned} \overline{\pi}_D(\xi, \Sigma) &:= \big((\xi_{[v]} - \Sigma_{[v\succ} \Sigma^{-1}_{\prec v \succ} \xi_{\prec v \succ}, \Sigma_{[v\succ} \Sigma^{-1}_{\prec v\succ}, \Sigma_{[v]\bullet}) \bigm| v \in V \big) \\ &\in \times \big(\mathbf{M}([v] \times N) \times \mathbf{M}([v] \times \prec v \succ) \times \mathbf{P}([v]) \bigm| v \in V \big) \end{aligned}$$

is called the family of *D-parameters* of (ξ, Σ).

Proposition 6.1. If L is a D-subspace, the mapping

$$\overline{\pi}_D \colon L \times \mathbf{P}(D; I) \to \times \big(L_{[v]} \times \mathbf{M}([v] \times \prec v \succ) \times \mathbf{P}([v]) \bigm| v \in V \big) \ =: \ \overline{\Pi}(L, D; I) \tag{6.11}$$

is bijective. Thus, every $(\xi, \Sigma) \in L \times \mathbf{P}(D; I)$ is uniquely determined by its D-parameters.

Proof. Since

$$\overline{\Pi}(L, D; I) = \big(\times (L_{[v]} | v \in V) \big) \times \Pi(D; I),$$

the inclusion

$$\overline{\pi}_D(L \times \mathbf{P}(D; I)) \subseteq \overline{\Pi}(L, D; I)$$

follows from (4.3) and conditions (i) and (iii). To see that $\overline{\pi}_D$ is injective, suppose that $\overline{\pi}_D(\xi, \Sigma) = \overline{\pi}_D(\xi', \Sigma')$ for $(\xi, \Sigma), (\xi', \Sigma') \in L \times \mathbf{P}(D; I)$. Then $\Sigma = \Sigma'$ by Proposition 4.1, hence

$$\xi_{[v]} - \Sigma_{[v\succ} \Sigma^{-1}_{\prec v\succ} \xi_{\prec v\succ} = \xi'_{[v]} - \Sigma_{[v\succ} \Sigma^{-1}_{\prec v\succ} \xi'_{\prec v\succ} \quad \forall v \in V. \tag{6.12}$$

Now choose a never-decreasing listing $v_1, \dots, v_r$ of V and apply (6.12) successively for $v = v_1, \dots, v_r$, together with (5.2), to obtain $\xi_1 = \xi'_1, \dots, \xi_r = \xi'_r$, hence $\xi = \xi'$.

To see that $\overline{\pi}_D$ is surjective, consider any $((\mu_v, \beta_v, \Lambda_v) \mid v \in V) \in \overline{\Pi}(L, D; I)$. Again select a never-decreasing listing $v_1, \dots, v_r$ of V and augment the general Step m of the Reconstruction Algorithm (cf. Section 5) with the following additional relation:

Step m' :
$$\xi_{[m]} = \mu_m + \beta_m \xi_{\prec m \succ}.$$

By (5.2) and condition (iii), $\xi_{[m]} \in L_{[m]}, m = 1, \dots, r$, so by condition (i) and Proposition 4.1, this Augmented Reconstruction Algorithm produces a pair $(\xi, \Sigma) \in L \times \mathbf{P}(D; I)$ such that $\pi_D(\Sigma) = \big((\beta_v, \Lambda_v) \bigm| v \in V \big)$. It follows from the relations in Step m', $m = 1, \dots, r$, that in fact $\overline{\pi}_D(\xi, \Sigma) = ((\mu_v, \beta_v, \Lambda_v) | v \in V)$, hence $\overline{\pi}_D$ is surjective.

7. Maximum likelihood estimation in a normal linear ADG model.

By Proposition 4.2, the likelihood function (LF) based on an observation y from the normal linear ADG model $\mathbf{N}_{I\times N}(L,D)$ in (6.9) has the following factorization:

$$
\begin{aligned}
&(L\times\mathbf{P}(D;I))\times\mathbf{M}(I\times N)\to\,]0,\infty[\\
(7.1)\qquad &\qquad ((\xi,\Sigma),y)\mapsto|\Sigma|^{-n/2}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\Sigma^{-1}(y-\xi)(y-\xi)^t\big)\Big\}
\end{aligned}
$$

$$
=\prod\Big(\,|\Sigma_{[v]\bullet}|^{-n/2}\times\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\Sigma_{[v]\bullet}^{-1}(y_{[v]}-\xi_{[v]}-\Sigma_{[v\succ}\Sigma_{\prec v\succ}^{-1}(y_{\prec v\succ}-\xi_{\prec v\succ}))(\cdots)^t\big)\Big\}\;\Big|\;v\in V\Big).
$$

For each $v\in V$, let $P_v:=P_{L_{[v]}}\in\mathbf{M}(N)$ denote the projection matrix corresponding to the MANOVA subspace $L_{[v]}$ and set $Q_v:=Q_{L_{[v]}}=1_N-P_v$. By the orthogonality of P_v and Q_v, the final expression in (7.1) has the following further factorization:

$$
\begin{aligned}
\prod\Big(\,|\Lambda_v|^{-n/2}\cdot\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\Lambda_v^{-1}(y_{[v]}P_v-\mu_v-\beta_v y_{\prec v\succ}P_v)(\cdots)^t\big)\Big\}\\
\cdot\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\Lambda_v^{-1}(y_{[v]}Q_v-\beta_v y_{\prec v\succ}Q_v)(\cdots)^t\big)\Big\}\;\Big|\;v\in V\Big),
\end{aligned}
$$

where $((\mu_v,\beta_v,\Lambda_v)|v\in V)\equiv\overline{\pi}(\xi,\Sigma)$ are the D-parameters of $(\xi,\Sigma)\in L\times\mathbf{P}(D;I)$. By Proposition 6.1, the parameter space $L\times\mathbf{P}(D;I)$ factors into the product of the ranges of the D-parameters.

It now follows readily from well-known results for the MANOVA model that the MLE $(\hat{\xi}(y),\hat{\Sigma}(y))$ of (ξ,Σ) is unique and exists for a.e. $y\in\mathbf{M}(I\times N)$ [Lebesgue] if and only if

$$
(7.2)\qquad n\geq\max\big\{\,p_v+|\preceq v\succeq|\;\big|\;v\in V\big\},
$$

where $p_v:=\mathrm{tr}(P_v)=\dim(K_v)$ with $K_v:=\mathrm{row}(P_v)$. In this case, the D-parameters $\big((\hat{\mu}_v,\hat{\beta}_v,\hat{\Lambda}_v)\;\big|\;v\in V\big)$ of the MLE $(\hat{\xi},\hat{\Sigma})$ are determined by the usual formulas for regression estimators:

$$
\begin{aligned}
\hat{\mu}_v&=y_{[v]}P_v-y_{[v]}Q_v y_{\prec v\succ}^t\big(y_{\prec v\succ}Q_v y_{\prec v\succ}^t\big)^{-1}y_{\prec v\succ}P_v\\
(7.3)\qquad\hat{\beta}_v&=y_{[v]}Q_v y_{\prec v\succ}^t\big(y_{\prec v\succ}Q_v y_{\prec v\succ}^t\big)^{-1}\\
n\hat{\Lambda}_v&=y_{[v]}Q_v y_{[v]}^t-y_{[v]}Q_v y_{\prec v\succ}^t\big(y_{\prec v\succ}Q_v y_{\prec v\succ}^t\big)^{-1}y_{\prec v\succ}Q_v y_{[v]}^t.
\end{aligned}
$$

The MLE $(\hat{\xi},\hat{\Sigma})$ itself may be reconstructed from these estimated D-parameters by means of the Augmented Reconstruction Algorithm described in the proof of Proposition 6.1.

Finally, when (7.2) holds, it follows from (4.8), (6.8), and the relation $\hat{\Lambda}_v=\hat{\Sigma}_{[v]\bullet}$ that the maximum of the LF in (7.1) is

$$
(7.4)\qquad\prod\big(\,|\hat{\Sigma}_{[v]\bullet}(y)|^{-n/2}\exp\{-n|[v]|/2\}\;\big|\;v\in V\big)=|\hat{\Sigma}(y)|^{-n/2}\exp\{-n|I|/2\}.
$$

Unlike the classical multivariate linear regression model, the normal linear ADG model $\mathbf{N}_{I\times N}(L,D)$ is a curved exponential family in general, so the MLE need not be a complete or sufficient statistic.

8. Distribution of the empirical generalized variance.

In this section we apply (4.8) to derive the distribution of the empirical generalized variance $|\hat{\Sigma}|$, where $\hat{\Sigma}\equiv\hat{\Sigma}(y)$ is the MLE of Σ, obtained implicitly in Section 7, based on an observation y from the normal linear ADG model $\mathbf{N}_{I\times N}(L,D)$. This distribution will be applied in Section 9 to obtain the central ($\equiv$ null) distribution of the likelihood ratio statistic for testing one normal linear ADG model against another.

Let $v_1,\ldots,v_r$ be a never-decreasing listing of the elements in V. As in Section 5, we denote $[v_m]$ by $[m]$, $\prec v_m \succ$ by $\prec m\succ$, and $\hat{\Lambda}_{v_m}$ by $\hat{\Lambda}_m$, $m=1,\ldots,r$. By the local Markov property, under the model $\mathbf{N}_{I\times N}(L,D)$ the conditional distribution of $y_{[r]}$ given $y_{[r-1]},\ldots,y_{[1]}$ is the same as the conditional distribution given $y_{\prec r\succ}$, hence by (7.3) the conditional distribution of $n\hat{\Lambda}_r\equiv n\hat{\Sigma}_{[r]\bullet}$ given $y_{[r-1]},\ldots,y_{[1]}$ is the same as the conditional distribution given $y_{\prec r\succ}$. By well-known results for the MANOVA model, this conditional distribution is the Wishart distribution $\mathcal{W}(\Sigma_{[r]\bullet},f_r)$ with $f_r:=n-p_r-|\prec r\succ|$ degrees of freedom and expectation $f_r\Sigma_{[r]\bullet}$, where $p_r:=\operatorname{tr}(P_r)$ with $P_r:=P_{v_r}$. Since this conditional distribution does not depend on $y_{[r-1]},\ldots,y_{[1]}$, $\hat{\Sigma}_{[r]\bullet}$ is independent of $(y_{[r-1]},\ldots,y_{[1]})$.

For $m=r-1,\ldots,1$, $\hat{\Sigma}_{[m]\bullet}$ depends on $y\in\mathbf{M}(I\times N)$ only through $y_{[m]},\ldots,y_{[1]}$. By repeating the preceding argument, we see that the conditional distribution of $n\hat{\Sigma}_{[m]\bullet}$ given $y_{[m-1]},\ldots,y_{[1]}$ is the same as its conditional distribution given $y_{\prec m\succ}$, i.e., the Wishart distribution $\mathcal{W}(\Sigma_{[m]\bullet},f_m)$ with $f_m:=n-p_m-|\prec m\succ|$ degrees of freedom and expectation $f_m\Sigma_{[m]\bullet}$, where $p_m:=tr(P_m)$ with $P_m:=P_{v_m}$. Thus $\hat{\Sigma}_{[r]\bullet},\ldots,\hat{\Sigma}_{[m]\bullet}$, and $(y_{[m-1]},\ldots,y_{[1]})$ are mutually independent.

We conclude that

$$\hat{\Sigma}_{[1]\bullet},\ldots,\hat{\Sigma}_{[r]\bullet} \text{ are mutually independent,} \tag{8.1}$$

$$n\hat{\Sigma}_{[m]\bullet}\ \sim\ \mathcal{W}(\Sigma_{[m]\bullet},f_m),\ m=1,\ldots,r. \tag{8.2}$$

In particular, it follows from (4.8) above and equation (15) in Anderson (1984), p. 264, that the α-th moment of the empirical generalized variance $|\hat{\Sigma}|$ is given by

$$\mathrm{E}(|\hat{\Sigma}|^\alpha)= \tag{8.3}$$

$$|\Sigma|^\alpha(2/n)^{\alpha|I|}\prod\Big(\prod\Big(\frac{\Gamma(\frac{1}{2}(n-p_v-|\prec v\succ|-i+1)+\alpha)}{\Gamma(\frac{1}{2}(n-p_v-|\prec v\succ|-i+1))}\ \Big|\ i=1,\ldots,|[v]|\Big)\ \Big|\ v\in V\Big).$$

9. Testing one normal linear ADG model against another.

In this section we address the general hypothesis-testing problem for normal linear ADG models. The general testing problem is formulated, the likelihood ratio (LR) test is derived (Proposition 9.1), and the null ($\equiv$ central) distribution of the LR statistic is specified in terms of its moments (Proposition 9.2). Examples are presented in §13.3.

Submodels $\mathbf{N}_{I\times N}(M, E)$ of a normal linear ADG model $\mathbf{N}_{I\times N}(L, D)$ are now introduced. As in Section 3, let $D \equiv (V, R)$ be an ADG such that $I = \dot\cup(I_v|v \in V)$ is a disjoint partitioning of the index set I, let $E \equiv (W, S)$ be a second ADG with an associated family $(I_w|w \in W)$ of disjoint subsets of I, and let $\psi: E \to D$ be a surjective ADG homomorphism such that

$$I_v = \dot\cup(\, I_w \mid w \in W, \psi(w) = v), \quad v \in V. \tag{9.1}$$

Thus also $I = \dot\cup(I_w|w \in W)$, so that (3.6) and (3.7) hold with $\mathbf{X} = \mathbf{R}^I$, $\mathbf{X}_v = \mathbf{R}^{I_v}$ and $\mathbf{Y}_w = \mathbf{R}^{I_w}$. By Proposition 3.1(i), $\mathbf{N}_I(E)$ is a submodel of $\mathbf{N}_I(D)$, or equivalently, $\mathbf{P}(E; I) \subseteq \mathbf{P}(D; I)$. If also ψ is a proper homomorphism, then it follows from a proof similar to that of Proposition 3.1(ii) that in fact $\mathbf{N}_I(E)$ is a proper submodel of $\mathbf{N}_I(D)$, i.e., that $\mathbf{P}(E; I) \subset \mathbf{P}(D; I)$.

For the remainder of this section, assume that L is a D-subspace of $\mathbf{M}(I\times N)$ and M is an E-subspace of $\mathbf{M}(I\times N)$ such that $M \subseteq L$. Then $\mathbf{N}_{I\times N}(M, E)$ is a submodel of $\mathbf{N}_{I\times N}(L, D)$, so we may consider the problem of testing $\mathbf{N}_{I\times N}(M, E)$ vs. $\mathbf{N}_{I\times N}(L, D)$. More specifically, based on an observation $y \sim \mathcal{N}_{I\times N}(\xi, \Sigma \otimes 1_N)$, we may test

$$H_{M,E} : (\xi, \Sigma) \in M \times \mathbf{P}(E; I) \quad vs. \quad H_{L,D} : (\xi, \Sigma) \in L \times \mathbf{P}(D; I). \tag{9.2}$$

Remark 9.1. It follows from Definition 6.2, (9.1), and the order-preserving property of ψ that every D-subspace L of $\mathbf{M}(I\times N)$ is also an E-subspace of $\mathbf{M}(I\times N)$, but the converse is not valid.[4] Therefore, the general problem (9.2) includes the following two testing problems as special cases:

$$H_{L,E} : (\xi, \Sigma) \in L \times \mathbf{P}(E; I) \quad vs. \quad H_{L,D} : (\xi, \Sigma) \in L \times \mathbf{P}(D; I), \tag{9.3}$$

$$H_{M,E} : (\xi, \Sigma) \in M \times \mathbf{P}(E; I) \quad vs. \quad H_{L,E} : (\xi, \Sigma) \in L \times \mathbf{P}(E; I). \tag{9.4}$$

In order to test

$$H_{M,D} : (\xi, \Sigma) \in M \times \mathbf{P}(D; I) \quad vs. \quad H_{L,D} : (\xi, \Sigma) \in L \times \mathbf{P}(D; I), \tag{9.5}$$

however, it must also be assumed that M is a D-subspace of $\mathbf{M}(I\times N)$.

Before presenting the LR statistic λ for the general testing problem (9.2), a warning about notation is needed. For $v \in V$ and $w \in W$, the reader is reminded that the subsets of I denoted by $[v]$, $\prec v \succ$, $\preceq v \succeq$ and by $[w]$, $\prec w \succ$, $\preceq w \succeq$ depend not only upon v and w, respectively, but also upon the partitionings $I = \dot\cup(I_v|v \in V)$ and $I = \dot\cup(I_w|w \in W)$

[4] For example, let $V := \{1, 2\}$, $I := \{1, 2\}$, $I_1 \equiv [1] := \{1\}$, $I_2 \equiv [2] := \{2\}$, $D = 1 \longrightarrow 2$, and $E = 1 \qquad 2$, so that $\mathrm{id}_{\{1,2\}}: E \to D$ is a (proper) surjective ADG homomorphism satisfying (9.1). Then any subspace $L \subseteq \mathbf{M}(I\times N)$ of the form $L = L_{[1]} \times L_{[2]}$, where $L_{[i]} = \mathbf{R}^{[i]} \otimes K_i$, $i = 1, 2$ (cf. (6.2)), is an E-subspace, but L is a D-subspace iff $K_1 \subseteq K_2$.

and relations R and S associated with the ADGs D and E, respectively. Similarly, the projections $P_v := P_{L_{[v]}}$ and $P_w := P_{M_{[w]}}$ depend not only on v and w but also on the D-subspace L and the E-subspace M, respectively, hence so do their traces p_v and p_w.

We denote the MLEs of Σ under $H_{M,E}$ and $H_{L,D}$ by $\hat{\Sigma}_0$ and $\hat{\Sigma}$, respectively.

Proposition 9.1. Suppose that (7.2) holds, i.e., $n \geq \max\{\, p_v + |\preceq v \succeq| \;\big|\; v \in V\}$. Then $\hat{\Sigma}_0(y)$ and $\hat{\Sigma}(y)$ exist and are unique for a.e. $y \in \mathbf{M}(I \times N)$ [Lebesgue] under $H_{L,D}$, hence also under $H_{M,E}$. In this case the LR statistic $\lambda \equiv \lambda(y)$ for testing $H_{M,E}$ vs. $H_{L,D}$ exists for a.e. y and is given by

$$\lambda^{2/n} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} = \frac{\prod\big(\,|\hat{\Sigma}_{[v]\bullet}| \;\big|\; v \in V\big)}{\prod\big(\,|\hat{\Sigma}_{0[w]\bullet}| \;\big|\; w \in W\big)}. \tag{9.6}$$

Proof. The existence and uniqueness a.e. of $\hat{\Sigma}$ follows from (7.2). For $w \in W$ we shall show the following:

$$\preceq w \succeq \;\; \subseteq \;\; \preceq \psi(w) \succeq, \tag{9.7}$$

$$P_w \;=\; P_{\psi(w)} P_w; \tag{9.8}$$

these two relations imply that $|\preceq w \succeq| \;\leq\; |\preceq \psi(w) \succeq|$ and $p_w \leq p_{\psi(w)}$, respectively. Therefore

$$\begin{aligned} n &\geq \max\{\, p_v + |\preceq v \succeq| \;\big|\; v \in V\} \\ &= \max\{\, p_{\psi(w)} + |\preceq \psi(w) \succeq| \;\big|\; w \in W\} \\ &\geq \max\{\, p_w + |\preceq w \succeq| \;\big|\; w \in W\}, \end{aligned}$$

implying the existence and uniqueness a.e. of $\hat{\Sigma}_0$, also by (7.2). The two expressions for λ follow from (7.4).

To establish (9.7), first note that (9.1) implies that

$$I_w \subseteq I_{\psi(w)}, \quad w \in W, \tag{9.9}$$

while the order-preserving property of ψ implies that

$$\psi\big(\{w\} \dot{\cup} \mathrm{pa}_E(w)\big) \;\subseteq\; \{\psi(w)\} \dot{\cup} \mathrm{pa}_D(\psi(w)). \tag{9.10}$$

Thus

$$\preceq w \succeq \;\equiv\; I_{\{w\} \dot{\cup} \mathrm{pa}_E(w)} \;\subseteq\; I_{\psi\left(\{w\} \dot{\cup} \mathrm{pa}_E(w)\right)} \;\subseteq\; I_{\{\psi(w)\} \dot{\cup} \mathrm{pa}_D(\psi(w))} \;\equiv\; \preceq \psi(w) \succeq,$$

which yields (9.7). Next, (9.9) implies that $[w] \subseteq [\psi(w)]$; since $M \subseteq L$ this in turn implies that $M_{[w]} \subseteq L_{[\psi(w)]}$, from which (9.8) follows.

Proposition 9.2. Suppose that (7.2) holds. Under the null hypothesis $H_{M,E}$ in (9.2), the LR statistic λ and the MLEs $(\hat{\Sigma}_{0[w]\bullet}|w \in W)$ are mutually independent. Under $H_{M,E}$, the $2\alpha/n$-th moment of λ is given by

(9.11)
$$\mathrm{E}(\lambda^{2\alpha/n}) = \frac{\mathrm{E}(|\hat{\Sigma}|^{\alpha})}{\mathrm{E}(|\hat{\Sigma}_0|^{\alpha})} = \frac{\prod\Big(\prod\Big(\frac{\Gamma(\frac{1}{2}(n-p_v-|\preceq v\succeq|-i+1)+\alpha)}{\Gamma(\frac{1}{2}(n-p_v-|\preceq v\succeq|-i+1))}\;\Big|\; i=1,\cdots,|[v]|\Big)\;\Big|\; v\in V\Big)}{\prod\Big(\prod\Big(\frac{\Gamma(\frac{1}{2}(n-p_w-|\preceq w\succeq|-i+1)+\alpha)}{\Gamma(\frac{1}{2}(n-p_w-|\preceq w\succeq|-i+1))}\;\Big|\; i=1,\cdots,|[w]|\Big)\;\Big|\; w\in W\Big)}.$$

Since $\sum\big(|[v]| \;\big|\; v \in V\big) = \sum\big(|[w]| \;\big|\; w \in W\big)$, these moments can be used to obtain the Box approximation for the central distribution of $-2\log\lambda$; see Anderson (1984, pp. 311-316) and [AP] (1995a, pp. 25-26). (The approximation given by Ledet Jensen (1991) should be somewhat more accurate.)

Remark 9.2. If we set $L = M = \{0\}$ in the general testing problem (9.2), then, since the LR statistic λ in (9.6) satisfies $0 \le \lambda \le 1$, we obtain the following extension of the classical Hadamard-Fischer determinantal inequality. Let $D \equiv (V, R)$ and $E \equiv (W, S)$ be two ADGs and let $\psi: E \to D$ be a surjective ADG homomorphism satisfying (9.1), so that $I = \dot{\cup}(I_v|v \in V) = \dot{\cup}(I_w|w \in W)$. Then (recall Footnote 3) for any positive definite matrix $\Sigma \in \mathbf{P}(I)$,

$$\prod\Big(\frac{|\Sigma_{\preceq v\succeq}|}{|\Sigma_{\prec v\succ}|}\;\Big|\; v\in V\Big) \;\le\; \prod\Big(\frac{|\Sigma_{\preceq w\succeq}|}{|\Sigma_{\prec w\succ}|}\;\Big|\; w\in W\Big). \tag{9.12}$$

In particular, by taking D to be the trivial ADG with only one vertex, this reduces to the inequality

$$|\Sigma| \;\le\; \prod\Big(\frac{|\Sigma_{\preceq w\succeq}|}{|\Sigma_{\prec w\succ}|}\;\Big|\; w\in W\Big). \tag{9.13}$$

Remark 9.3. Since the class of normal LCI models is a subset of the class of normal ADG models (see Remark 4.1), the results in this section may be regarded as extensions of those in [AP] (1995a) where, furthermore, no non-zero mean-value subspaces were considered. These results also extend results concerning testing one decomposable covariance selection model against another (cf. Porteous (1989), Andersen *et al* (1995, §7.6.1), Eriksen (1996), Lauritzen (1996)) where, again, general mean-value subspaces were not considered.

Proof of Proposition 9.2. To begin the derivation of the central distribution of λ, choose a never-decreasing listing $v_1, \ldots. v_r$ of V. For each $m = 1, \ldots, r$, let $w_{m1}, \ldots, w_{mq_m}$ be a never-decreasing listing of $\psi^{-1}(v_m)$. It follows from (9.1) and the surjective and order-preserving properties of ψ that the combined sequence $w_{11}, \ldots, w_{1q_1}, \ldots, w_{r1}, \ldots, w_{rq_r}$ is a never-decreasing listing of the members of W, which in turn implies that

$$\begin{aligned}\mathrm{pa}_E(w_{mk}) &\subseteq \{\, w_{ij} \mid i = 1, \ldots, m-1,\; j = 1, \ldots, q_i \,\}\dot{\cup}\{\, w_{mj} \mid j = 1, \ldots, k-1 \,\} \\ &\subseteq \mathrm{nd}_E(w_{mk}).\end{aligned} \tag{9.14}$$

For each $m = 1, \ldots, r$ and $k = 1, \ldots, q_m$, define

$$\langle w_{mk} \rangle := I_{\{w_{ij} | i=1,\ldots,m-1,\ j=1,\ldots,q_i\} \dot\cup \{w_{mj} | j=1,\ldots,k-1\}}.$$

For notational convenience, we usually abbreviate v_m, $[v_m]$, $\prec v_m \succ$, $[v_m \succ$ by m, $[m]$, $\prec m \succ$, $[m \succ$, respectively, and w_{mk}, $[w_{mk}]$, $\prec w_{mk} \succ$, $[w_{mk} \succ$, $\langle w_{mk} \rangle$ by mk, $[mk]$, $\prec mk \succ$, $[mk \succ$, $\langle mk \rangle$, respectively.

From (9.6), λ can be expressed as follows:

$$\lambda^{2/n} = \prod(\eta_m \mid m = 1, \ldots, r), \tag{9.15}$$

where

$$\eta_m := \frac{|\hat\Sigma_{[m]\bullet}|}{\prod\big(|\hat\Sigma_{0[mk]\bullet}| \mid k = 1, \ldots, q_m \big)}. \tag{9.16}$$

Furthermore, by eqn. (2.21) of [AP] (1995b), for $m = 1, \ldots, r$ we have that

$$\eta_m = \prod(\omega_{mk} \mid k = 1, \ldots, q_m), \tag{9.17}$$

where

$$\omega_{mk} := \frac{\big|\hat\Sigma_{[mk]\bullet(\prec m \succ \dot\cup [m1] \dot\cup \cdots \dot\cup [m(k-1)])}\big|}{|\hat\Sigma_{0[mk]\bullet}|}. \tag{9.18}$$

For each $m = 1, \ldots, r$ and $k = 1, \ldots, q_m$ and for each fixed $y_{\langle mk \rangle} \in \mathbf{M}(\langle mk \rangle \times N)$, we shall show that ω_{mk} has the form of Wilks' LR criterion for testing one MANOVA model against another (cf. (9.27)), and thereby derive the central distribution of λ.

By (9.14), under the hypothesis $H_{M,E}$ the conditional distribution of $y_{[mk]}$ given $y_{\langle mk \rangle}$ is as follows:

$$y_{[mk]} \mid y_{\langle mk \rangle} \sim \mathcal{N}_{[mk] \times N}(\mu_{mk} + \beta_{mk} y_{\prec mk \succ_E}, \Lambda_{mk} \otimes 1_N), \tag{9.19}$$

(recall Footnote 3), where

$$\begin{aligned} \mu_{mk} &:= \xi_{[mk]} - \Sigma_{[mk \succ_E} \Sigma^{-1}_{\prec mk \succ_E} \xi_{\prec mk \succ_E} && \in M_{[mk]}, \\ \beta_{mk} &:= \Sigma_{[mk \succ_E} \Sigma^{-1}_{\prec mk \succ_E} && \in \mathbf{M}([mk] \times \prec mk \succ_E), \\ \Lambda_{mk} &:= \Sigma_{[mk]\bullet \prec mk \succ_E} && \in \mathbf{P}([mk]). \end{aligned} \tag{9.20}$$

It follows from Proposition 6.1 with L, D replaced by M, E that the range of the parameter $(\mu_{mk}, \beta_{mk}, \Lambda_{mk})$ under $H_{M,E}$ is

$$M_{[mk]} \times \mathbf{M}([mk] \times \prec mk \succ_E) \times \mathbf{P}([mk]).$$

Therefore, the range of $(\mu_{mk} + \beta_{mk} y_{\prec mk \succ_E}, \Lambda_{mk})$ under $H_{M,E}$ is

$$M(y_{\prec mk \succ_E}) \times \mathbf{P}([mk]),$$

where, for a.e. $y_{\prec mk \succ_E} \in \mathbf{M}(\prec mk \succ_E \times N)$, $M(y_{\prec mk \succ_E}) \subseteq \mathbf{M}([mk] \times N)$ is the MANOVA subspace[5]

$$M(y_{\prec mk \succ_E}) := M_{[mk]} \oplus \{ \beta_{mk} y_{\prec mk \succ_E} \mid \beta_{mk} \in \mathbf{M}([mk] \times \prec mk \succ_E) \}. \tag{9.21}$$

Next, let $\tilde{E} \equiv (W, \tilde{S})$ be the ADG constructed from E and D in the proof of Proposition 3.1; recall from (3.8) that

$$\prec mk \succ_{\tilde{E}} = \prec m \succ \dot{\cup} [m1] \dot{\cup} \cdots \dot{\cup} [m(k-1)] \subseteq \langle mk \rangle. \tag{9.22}$$

By (3.4) and (3.2), under the hypothesis $H_{L,D}$ the conditional distribution of $y_{[mk]}$ given $y_{\langle mk \rangle}$ is the following (recall Footnote 3):

$$y_{[mk]} \mid y_{\langle mk \rangle} \sim \mathcal{N}_{[mk] \times N}(\mu_{mk} + \beta_{mk} y_{\prec mk \succ_{\tilde{E}}}, \Lambda_{mk} \otimes 1_N), \tag{9.23}$$

where now

$$\begin{aligned} \mu_{mk} &:= \xi_{[mk]} - \Sigma_{[mk \succ_{\tilde{E}}} \Sigma^{-1}_{\prec mk \succ_{\tilde{E}}} \xi_{\prec mk \succ_{\tilde{E}}} && \in L_{[mk]} \text{ (see below)}, \\ \beta_{mk} &:= \Sigma_{[mk \succ_{\tilde{E}}} \Sigma^{-1}_{\prec mk \succ_{\tilde{E}}} && \in \mathbf{M}([mk] \times \prec mk \succ_{\tilde{E}}), \\ \Lambda_{mk} &:= \Sigma_{[mk] \bullet \prec mk \succ_{\tilde{E}}} && \in \mathbf{P}([mk]). \end{aligned} \tag{9.24}$$

It is easily verified that $\psi \colon \tilde{E} \to D$ is also a surjective ADG homomorphism, so by Remark 9.1 with E replaced by $\tilde{E}$, L is also an $\tilde{E}$-subspace. Since $\xi \in L$ under $H_{L,D}$, the relation $\mu_{mk} \in L_{[mk]}$ holds.

It follows from Proposition 6.1 with D replaced by $\tilde{E}$ that the range of the parameter $(\mu_{mk}, \beta_{mk}, \Lambda_{mk})$ under $H_{L,D}$ is

$$L_{[mk]} \times \mathbf{M}([mk] \times \prec mk \succ_{\tilde{E}}) \times \mathbf{P}([mk]).$$

Therefore, the range of $(\mu_{mk} + \beta_{mk} y_{\prec mk \succ_{\tilde{E}}}, \Lambda_{mk})$ under $H_{L,D}$ is

$$L(y_{\prec mk \succ_{\tilde{E}}}) \times \mathbf{P}([mk]),$$

where, for a.e. $y_{\prec mk \succ_{\tilde{E}}} \in \mathbf{M}(\prec mk \succ_{\tilde{E}} \times N)$, $L(y_{\prec mk \succ_{\tilde{E}}}) \subseteq \mathbf{M}([mk] \times N)$ is the MANOVA subspace

$$L(y_{\prec mk \succ_{\tilde{E}}}) := L_{[mk]} \oplus \{ \beta_{mk} y_{\prec mk \succ_{\tilde{E}}} \mid \beta_{mk} \in \mathbf{M}([mk] \times \prec mk \succ_{\tilde{E}}) \}. \tag{9.25}$$

[5] The occurrence of the direct sum in (9.21) for a.e. $y_{\prec mk \succ_E}$ follows, for example, from Theorem 2.3 of Eaton and Perlman (1973); the same holds for the direct sum in (9.25).

Note now that ω_{mk} in (9.18) can be expressed as

$$\omega_{mk} = \frac{|\hat{\Sigma}_{[mk]\bullet\prec mk\succ_{\tilde{E}}}|}{|\hat{\Sigma}_{0[mk]\bullet\prec mk\succ_E}|}. \tag{9.26}$$

By (3.5) with $\breve{D}, D$ replaced by $E, \tilde{E}$, $\mathrm{pa}_E(mk) \subseteq \mathrm{pa}_{\tilde{E}}(mk)$, hence $\prec mk\succ_E \subseteq \prec mk\succ_{\tilde{E}}$. Also, $M \subseteq L$ implies that $M_{[mk]} \subseteq L_{[mk]}$, hence $M(y_{\prec mk\succ_E}) \subseteq L(y_{\prec mk\succ_{\tilde{E}}})$. Thus for each fixed $y_{\langle mk\rangle} \in \mathbf{M}(\langle mk\rangle \times N)$, we may consider the conditional problem of testing the normal MANOVA models (cf. (6.5))

$$\mathbf{N}_{[mk]\times N}(M(y_{\prec mk\succ_E})) \quad vs. \quad \mathbf{N}_{[mk]\times N}(L(y_{\prec mk\succ_{\tilde{E}}})). \tag{9.27}$$

It follows from (6.8) that the LR statistic $\lambda_{mk} \equiv \lambda_{mk}(y_{\langle mk\rangle})$ for this conditional testing problem is given by

$$\lambda_{mk}^{2/n} = \frac{|\hat{\Lambda}_{mk}|}{|\hat{\Lambda}_{0mk}|}, \tag{9.28}$$

where $\hat{\Lambda}_{0mk}$ and $\hat{\Lambda}_{mk}$ are the respective MLEs of Λ_{mk} under the conditional models $\mathbf{N}_{[mk]\times N}(M(y_{\prec mk\succ_E}))$ ($\equiv$ (9.19)) and $\mathbf{N}_{[mk]\times N}(L(y_{\prec mk\succ_{\tilde{E}}}))$ ($\equiv$ (9.23)). Straightforward but lengthy algebra using (6.7) and the relation $Q_{L_{[mk]}} = Q_{L_m}$ shows, however, that

$$\begin{aligned} \hat{\Lambda}_{0mk} &= \hat{\Sigma}_{0[mk]\bullet\prec mk\succ_E}, \\ \hat{\Lambda}_{mk} &= \hat{\Sigma}_{[mk]\bullet\prec mk\succ_{\tilde{E}}}. \end{aligned} \tag{9.29}$$

We conclude that for each fixed $y_{\langle mk\rangle}$, $\omega_{mk} = \lambda_{mk}^{2/n}$, as asserted after (9.18).

Therefore, under the conditional MANOVA model $\mathbf{N}_{[mk]\times N}(M(y_{\prec mk\succ_E}))$ ($\equiv$ (9.19)) in (9.27), for each fixed $y_{\langle mk\rangle}$ we have

$$\begin{aligned} \omega_{mk} \mid y_{\langle mk\rangle} &\sim U_{|[mk]|,\,p_m - p_{0mk} + |\prec mk\succ_{\tilde{E}}| - |\prec mk\succ_E|,\,n - p_m + |\prec mk\succ_{\tilde{E}}|} \\ n\hat{\Lambda}_{0mk} \equiv n\hat{\Sigma}_{0[mk]\bullet} \mid y_{\langle mk\rangle} &\sim \mathcal{W}(\Lambda_{mk} \equiv \Sigma_{[mk]\bullet}, n - p_{0mk} + |\prec mk\succ_E|), \end{aligned} \tag{9.30}$$

where $U_{a,b,c}$ denotes Wilk's U distribution (Anderson (1984, §8.4)) and

$$\begin{aligned} p_m &:= \mathrm{tr}(P_{L_{[m]}}) \equiv \mathrm{tr}(P_{L_{[mk]}}) \\ p_{0mk} &:= \mathrm{tr}(P_{M_{[mk]}}) \\ \hat{\Sigma}_{0[mk]\bullet} &:= \hat{\Sigma}_{0[mk]\bullet\prec mk\succ_E} \\ \Sigma_{[mk]\bullet} &:= \Sigma_{[mk]\bullet\prec mk\succ_E}. \end{aligned}$$

Thus ω_{mk} is a (conditionally) ancillary statistic and the MLE $\hat{\Lambda}_{0mk} \equiv \hat{\Sigma}_{0[mk]\bullet}$ is a function of the (conditionally) complete sufficient statistic (cf. (6.7)), so by Basu's Lemma, ω_{mk} and $\hat{\Sigma}_{0[mk]\bullet}$ are (conditionally) independent. With (9.30), this implies that under $H_{M,E}$,

$$\begin{aligned} \omega_{mk} &\perp\!\!\!\perp \hat{\Sigma}_{0[mk]\bullet} \mid y_{\langle mk\rangle}, \\ \omega_{mk} &\perp\!\!\!\perp y_{\langle mk\rangle}, \\ \hat{\Sigma}_{0[mk]\bullet} &\perp\!\!\!\perp y_{\langle mk\rangle}, \end{aligned} \tag{9.31}$$

hence by (3.2),

$$\omega_{mk} \perp\!\!\!\perp \hat{\Sigma}_{0[mk]\bullet} \perp\!\!\!\perp y_{\langle mk \rangle}. \tag{9.32}$$

Since $\hat{\Lambda}_{m(k-1)}$ and $\hat{\Lambda}_{0m(k-1)}$ are functions of $y_{\langle mk \rangle}$, so are $\omega_{m(k-1)}$ and $\hat{\Sigma}_{0[m(k-1)]\bullet}$. Thus we may use an inductive argument to conclude that under the null hypothesis $H_{M,E}$, the $2|W|$ statistics $\big((\omega_{mk}, \hat{\Sigma}_{0[mk]\bullet}) \bigm| m = 1,\dots,r,\ k = 1,\dots,q_m\big)$ are mutually independent, and their unconditional distributions are the same as their conditional distributions in (9.27). Therefore, by (9.15) - (9.18), the LR statistic λ and the MLEs $(\hat{\Sigma}_{0[w]\bullet}|w \in W)$ are mutually independent under $H_{M,E}$, as asserted.

Finally, it follows from (9.6) that

$$|\hat{\Sigma}| = \lambda^{2/n}|\hat{\Sigma}_0| = \lambda^{2/n} \prod\big(|\hat{\Sigma}_{0[w]\bullet}| \bigm| w \in W\big),$$

so by the independence just established, under $H_{M,E}$ the $2\alpha/n$-th moment of λ is given by

$$\mathrm{E}(\lambda^{2\alpha/n}) = \frac{\mathrm{E}(|\hat{\Sigma}|^\alpha)}{\mathrm{E}(|\hat{\Sigma}_0|^\alpha)}.$$

Because $H_{M,E} \subseteq H_{L,D}$, the second equality in (9.11) follows from (8.3).

10. Two characterizations of D-linear subspaces.

In this section we study further the structure of D-subspaces. As before, $D \equiv (V, R)$ is an ADG and $I = \dot{\cup}(I_v|v \in V)$ is an associated partition of the finite index set I with each $I_v \neq \emptyset$.

To begin, note that in Definition 6.2, condition (iii) is equivalent to the following:

(iv) $\forall u, v \in V$ with $u <_D v$, $\mathbf{M}([v]\times[u])L_{[u]} \subseteq L_{[v]}$, or equivalently, by (6.2), $K_{L_{[u]}} \subseteq K_{L_{[v]}}$.

Thus the definition of a D-subspace depends on the relation $\prec_D$ only through the induced transitive relation $<_D$. It follows that *L is a D-subspace if and only if L is a $T(D)$-subspace,* where the ADG $T(D)$ is the *transitive closure* of D. That is, $T(D)$ is the ADG with vertex set V and with transitive binary relation $\prec_{T(D)}$ defined as follows: for $u, v \in V$, $u \prec_{T(D)} v$ iff $u <_D v$.

Remark 10.1. Since $T(D)$ is a transitive ADG ($\equiv$ TADG) or, equivalently, a partially ordered set ($\equiv$ poset), and since the ancestral rings $\mathbf{A}(D)$ and $\mathbf{A}(T(D))$ coincide, the fundamental Birkhoff duality between finite posets and finite distributive lattices (cf. Davey and Priestley (1990), Chapter 8, or [A] (1990), Theorem 3.2) can be applied as in [AMPT] (1995, 1997) to deduce that *L is a D-subspace of $\mathbf{M}(I\times N)$ if and only if L is a $\mathcal{K}(D)$-subspace of $\mathbf{M}(I\times N)$*. Here $\mathcal{K}(D) := \{I_A|A \in \mathbf{A}(D)\}$ is a ring of subsets (hence a finite distributive lattice) of I isomorphic to the ancestral ring $\mathbf{A}(D)$, and for any ring $\mathcal{K}$ of subsets of I, the relevant definition of a $\mathcal{K}$-subspace appears in Theorem 4.2 of [AP] (1994). We conclude that although the class of covariance models determined by the class of normal

ADG models is strictly larger than that determined by the class of normal LCI models, the class of linear regression subspaces naturally associated with normal ADG models coincides with that associated with normal LCI models.

Next, we present an algebraic characterization of D-subspaces in terms of their invariance under a linear class $\mathbf{M}(D;I)$ of generalized block-triangular $I\times I$ matrices determined by D (Proposition 10.1). This characterization can be used to verify that a specified regression subspace is a D-subspace - see Section 13.

For any $A \in \mathbf{M}(I)$ and $u, v \in V$, let $A_{[uv]}$ denote the $[u]\times[v]$ submatrix of A and define $A_{[v]} := A_{[vv]}$. Each $A \in \mathbf{M}(I)$ can be partitioned according to the decomposition $I = \dot\cup(\,[v] \mid v \in V)$ as follows:

$$A = (\,A_{[uv]} \mid u, v \in V).$$

Define

$$\text{(10.1)} \qquad \mathbf{M}(D;I) := \{\,A \in \mathbf{M}(I) \mid \forall u, v \in V,\ v \not\preceq u \Rightarrow A_{[uv]} = 0\,\},$$

$$\text{(10.2)} \qquad \mathbf{M}_1(D;I) := \{\,A \in \mathbf{M}(D;I) \mid \forall v \in V,\ A_{[v]} = 1_{[v]}\,\}.$$

For $v \in V$ and $A \in \mathbf{M}(I)$, partition $A_{\preceq v\succeq} \in \mathbf{M}(\preceq v\succeq \times \preceq v\succeq)$ according to the decomposition $\preceq v\succeq = \prec v\succ \dot\cup [v]$:

$$A_{\preceq v\succeq} = \begin{pmatrix} A_{\prec v\succ} & A_{\prec v]} \\ A_{[v\succ} & A_{[v]} \end{pmatrix},$$

where $A_{[v]} \in \mathbf{M}([v])$, $A_{[v\succ} \in \mathbf{M}([v]\times \prec v\succ)$, $A_{\prec v]} \in \mathbf{M}(\prec v] \times [v])$, $A_{\prec v\succ} \in \mathbf{M}(\prec v\succ \times \prec v\succ)$. By (10.1), if $A \in \mathbf{M}(D;I)$ then $A_{\prec v]} = 0\ \forall v \in V$. Furthermore, the following linear mapping is bijective:

$$\text{(10.3)} \qquad \begin{aligned} \mathbf{M}(D;I) &\to \times\big(\mathbf{M}([v]\times \prec v\succ) \times \mathbf{M}([v]) \;\big|\; v \in V\big) \\ A &\mapsto \quad \big((A_{[v\succ}, A_{[v]}) \;\big|\; v \in V\big). \end{aligned}$$

Proposition 10.1. A subspace $L \subseteq \mathbf{M}(I\times N)$ is a D-subspace if and only if

$$\text{(10.4)} \qquad \mathbf{M}(D;I)L \subseteq L.$$

Proof. For each $v \in V$ define

$$\widetilde{\mathbf{M}}([v]\times \prec v\succ) := \big\{\,A \equiv (A_{[tu]} \mid t, u \in V) \in \mathbf{M}(I) \;\big|\; A_{[tu]} = 0 \text{ unless } t = v \text{ and } u \prec v\big\},$$
$$\widetilde{\mathbf{M}}([v]) := \big\{\,A \equiv (A_{[tu]} \mid t, u \in V) \in \mathbf{M}(I) \;\big|\; A_{[tu]} = 0 \text{ unless } t = u = v\big\}.$$

Since the mapping (10.3) is bijective,

$$\text{(10.5)} \qquad \mathbf{M}(D;I) = \oplus\big(\widetilde{\mathbf{M}}([v]\times \prec v\succ) \oplus \widetilde{\mathbf{M}}([v]) \;\big|\; v \in V\big),$$

hence (10.4) holds iff both of the following two conditions hold for every $v \in V$:

$$\widetilde{\mathbf{M}}([v])L \subseteq L, \tag{10.6}$$

$$\widetilde{\mathbf{M}}([v]\times \prec v \succ)L \subseteq L. \tag{10.7}$$

It is straightforward to show that conditions (i) and (ii) of Definition 6.2 are together equivalent to (10.6), and that when (i) and (ii) hold, condition (iii) is equivalent to (10.7). This completes the proof.

It follows from Proposition 10.1 that the set of D-subspaces is closed under the operations of intersection and summation. The application of Proposition 10.1 to identify D-subspaces is illustrated in Section 13.

Remark 10.2. Clearly, $\mathbf{M}(D; I)$ contains 1_I and is a linear space, i.e., closed under addition and scalar multiplication, but is not necessarily closed under matrix multiplication, i.e., $\mathbf{M}(D; I)$ is not necessarily a matrix algebra. The matrix algebra generated by $\mathbf{M}(D; I)$, denoted by $\overline{\mathbf{M}}(D; I)$, is the set of all finite sums of finite products of matrices in $\mathbf{M}(D; I)$. It is easy to see that condition (10.4) is equivalent to the condition

$$\overline{\mathbf{M}}(D; I)L \subseteq L. \tag{10.8}$$

Lemma 10.1. The following four conditions are equivalent:

(i) D is transitive.

(ii) $\mathbf{M}(D; I)$ is closed under matrix multiplication, i.e., $\mathbf{M}(D; I)$ is a matrix algebra;

(iii) $\mathbf{M}(D; I)$ is closed under matrix inversion;

(iv) $\mathbf{M}(D; I)$ is closed under Jordan multiplication: $A, B \in \mathbf{M}(D; I) \Rightarrow AB+BA \in \mathbf{M}(D; I)$.

Proof. (i) $\Rightarrow$ (ii): If D is transitive, it follows from (10.1) and the standard relation

$$(AB)_{[uv]} = \sum (A_{[uw]} B_{[wv]} \mid w \in V), \quad A, B \in \mathbf{M}(I), \ u, v \in V, \tag{10.9}$$

that $\mathbf{M}(D; I)$ is closed under matrix multiplication.

(ii) $\Rightarrow$ (iii): (ii) implies that each nonsingular $A \in \mathbf{M}(D; I)$ determines an injective linear mapping $\psi_A : \mathbf{M}(D; I) \to \mathbf{M}(D; I)$ defined by $\psi_A(B) = AB$. A standard dimensionality argument shows that this mapping is also surjective, hence, since $1_I \in \mathbf{M}(D; I)$, $B := A^{-1} \in \mathbf{M}(D; I)$.

(iii) $\Rightarrow$ (iv): For $A \in \mathbf{M}(I)$ we have

$$\lim_{t \to 0} \frac{(1_I - tA)^{-1} - 1_I - tA}{t^2} = A^2. \tag{10.10}$$

By (iii), if $A \in \mathbf{M}(D; I)$ then the limit in (10.10) also is an element of $\mathbf{M}(D; I)$, since the latter is closed in $\mathbf{M}(I)$, hence $A^2 \in \mathbf{M}(D; I)$. Thus, $A, B \in \mathbf{M}(D; I) \Rightarrow A^2, B^2, (A+B)^2 \in \mathbf{M}(D; I) \Rightarrow AB + BA \in \mathbf{M}(D; I)$.

(iv) $\Rightarrow$ (i): If D is not transitive, then there exist $v, w, u \in V$ such that $v \prec w \prec u$ but $v \not\preceq u$. Define $A \equiv (\, A_{[u'v']} \mid u', v' \in V)$ and $B \equiv (\, B_{[w'v']} \mid w', v' \in V)$ as follows:

$$A_{[u'w']} := \begin{cases} 0, & \text{if } (u', w') \neq (u, w), \\ \tilde{A}, & \text{if } (u', w') = (u, w), \end{cases}$$

$$B_{[w'v']} := \begin{cases} 0, & \text{if } (w', v') \neq (w, v), \\ \tilde{B}, & \text{if } (w', v') = (w, v), \end{cases}$$

where $\tilde{A} \in \mathbf{M}([u] \times [w])$ and $\tilde{B} \in \mathbf{M}([w] \times [v])$ are chosen such that $\tilde{A}\tilde{B} \neq 0$. Then by (10.1) and (10.9), $A, B \in \mathbf{M}(D; I)$, $BA = 0$, but $AB \notin \mathbf{M}(D; I)$ since $(AB)_{[uv]} = \tilde{A}\tilde{B}$, hence $AB + BA \notin \mathbf{M}(D; I)$.

Lemma 10.2. $\overline{\mathbf{M}}(D; I) = \mathbf{M}(T(D); I)$.

Proof. From the definitions of $\mathbf{M}(D; I)$ and $T(D)$, $\mathbf{M}(D; I) \subseteq \mathbf{M}(T(D); I)$. By Lemma 10.1 applied to $T(D)$, $\mathbf{M}(T(D); I)$ is an algebra, hence $\overline{\mathbf{M}}(D; I) \subseteq \mathbf{M}(T(D); I)$. Next we establish the opposite inclusion. Since

$$\mathbf{M}(T(D); I) = \text{span}(E^{ij} \mid i \in [u], j \in [v], v \leq u),$$

where

$$E^{ij} := (\delta_{ii'}\delta_{jj'} \mid i', j' \in I) \in \mathbf{M}(I),$$

it suffices to show that $E^{ij} \in \overline{\mathbf{M}}(D; I)$ if $i \in [u], j \in [v], v \leq u$. In this case, $v \equiv w_k \prec \cdots \prec w_0 \equiv u$ for some $w_0, \ldots, w_k \in V$, $k \geq 0$. If $k \leq 1$ then $E^{ij} \in \mathbf{M}(D; I)$. If $k \geq 2$, then

$$E^{ij} = E^{l_0 l_1} \cdots E^{l_{k-1} l_k},$$

where $l_0 := i$, $l_k := j$, and l_ν is chosen arbitrarily in $[w_\nu]$, $\nu = 1, \ldots, k-1$. Since $E^{l_{\nu-1} l_\nu} \in \mathbf{M}(D; I)$ for $\nu = 1, \ldots, k$, $E^{ij} \in \overline{\mathbf{M}}(D; I)$ as required.

From Proposition 10.1, Remark 10.2, and Lemma 10.2, we may again conclude that L is a D-subspace iff L is a $T(D)$-subspace.

Remark 10.3. As in Remark 10.1, it can be seen that $\mathbf{M}(T(D); I) = \mathbf{M}(\mathcal{K}(D))$, where, for any ring $\mathcal{K}$, $\mathbf{M}(\mathcal{K})$ is defined in [AP] (1993, 1994).

11. Normal linear ADG models and block-recursive normal linear systems.

Consider again the normal linear ADG model $\mathbf{N}_{I \times N}(L, D)$ in (6.9), where D is an ADG as in Section 10 and $L \subseteq \mathbf{R}^{I \times N}$ is a D-subspace. An alternative interpretation of this model can be obtained by means of the following block-triangular decomposition of the set of covariance matrices associated with $\mathbf{N}_I(D)$:

Proposition 11.1.

$$\mathbf{P}(D; I) = \{A^{-1}\Gamma(A^{-1})^t \mid A \in \mathbf{M}_1(D; I),\ \Gamma = \text{diag}(\Gamma_{[v]} | v \in V), \Gamma_{[v]} \in \mathbf{P}([v]), v \in V\}.$$

Proof. By Proposition 4.2, if $\Sigma \in \mathbf{P}(D;I)$ then $\Sigma^{-1} = A^t\Gamma^{-1}A$, where $A \in \mathbf{M}_1(D;I)$ is given by (recall (10.3)) $A_{[v\succ} = -\Sigma_{[v\succ}\Sigma^{-1}_{\prec v\succ}$, $v \in V$, and $\Gamma \equiv \mathrm{diag}(\Gamma_{[v]}|v \in V)$ is given by $\Gamma_{[v]} = \Sigma_{[v]\bullet}$, $v \in V$. Conversely, if $\Sigma = A^{-1}\Gamma(A^{-1})^t$ for some $A \in \mathbf{M}_1(D;I)$ and $\Gamma = \mathrm{diag}(\Gamma_{[v]}|v \in V), \Gamma_{[v]} \in \mathbf{P}([v]), v \in V$, then for every $x \in \mathbf{R}^I$,

$$\mathrm{tr}(\Sigma^{-1}xx^t) = \mathrm{tr}(\Gamma^{-1}Ax(Ax)^t) = \sum\Big(\mathrm{tr}\big(\Gamma^{-1}_{[v]}(x_{[v]} + A_{[v\succ}x_{\prec v\succ})(\cdots)^t\big) \;\Big|\; v \in V\Big).$$

By Propositions 4.1 and 4.2, this implies that Σ is the unique element in $\mathbf{P}(D;I)$ determined by $\Sigma_{[v\succ}\Sigma^{-1}_{\prec v\succ} = -A_{[v\succ}$ and $\Sigma_{[v]\bullet} = \Gamma_{[v]}$, $v \in V$. This completes the proof.

By Proposition 11.1, $y \equiv (y_{[v]}|v \in V) \in \mathbf{R}^{I\times N}$ is an observable stochastic variable from the model $\mathbf{N}_{I\times N}(L, D)$ iff

$$y = \xi + A^{-1}z \tag{11.1}$$

for some $A \in \mathbf{M}_1(D;I)$ and some $\xi \in L$, where $z \equiv (z_{[v]}|v \in V) \in \mathbf{R}^{I\times N}$ is an unobservable stochastic variable such that $z \sim \mathrm{N}(0, \Gamma \otimes 1_N)$ for some $\Gamma := \mathrm{diag}(\Gamma_{[v]}|v \in V)$, $\Gamma_{[v]} \in \mathbf{P}([v])$, $v \in V$. From Proposition 10.1, Remark 10.2, and the fact that any matrix algebra containing 1_I (thus $\overline{\mathbf{M}}(D;I)$) is closed under matrix inversion, it follows that this representation is equivalent to the relation

$$Ay = \mu + z \tag{11.2}$$

for some $A \in \mathbf{M}_1(D;I)$ and some $\mu \in L$. In turn, (11.2) is equivalent to the following*block-recursive normal linear system with block-recursive regression subspaces*:

$$y_{[v]} + A_{[v\succ}y_{\prec v\succ} = \mu_{[v]} + z_{[v]},\ v \in V, \tag{11.3}$$

for some $A_{[v\succ} \in \mathbf{M}([v]\times \prec v\succ), v \in V$, and some $(\mu_{[v]}|v \in V) \in L$. The recursive nature of this system is determined by the acyclic property of D through the definition of $y_{\prec v\succ}$ and through the D-subspace restriction on $(\mu_{[v]}|v \in V) \in L$. Thus, the normal linear ADG model $\mathbf{N}_{I\times N}(L, D)$ is equivalent to a block-recursive linear system (and conversely).

Recursive linear systems have appeared frequently in the statistics literature, for example Wermuth (1980), Kiiveri *et al* (1984), Wermuth (1992), [AP] (1993, Remark 3.5), Cox and Wermuth (1996), Lauritzen (1996), and even more often in the econometrics literature, e.g. Goldberger (1964), Bollen (1989). The normal linear ADG models comprise the special subclass of recursive linear systems where the regression structure is so adapted to the covariance structure that the model can be decomposed into a product of standard MANOVA models, permitting explicit estimates and tests.

The representation of $\mathbf{P}(D;I)$ in Proposition 11.1 is equivalent to the representation

$$\mathbf{P}(D;I)^{-1} = \{A^tA \mid A \in \mathbf{M}(D;I), A \text{ nonsingular}\}. \tag{11.5}$$

of $\mathbf{P}(D;I)^{-1}$. The following representation (11.6) of $\mathbf{P}(D;I)$ holds iff D is transitive. (Also see Remark 2.4 of [AP] (1993).)

Proposition 11.2.

(11.6) $$\mathbf{P}(D;I) = \{AA^t \mid A \in \mathbf{M}(D;I), A \text{ nonsingular}\}$$

if and only if D is transitive.

Proof. If D is transitive, then by Lemma 10.1 ((i) $\Rightarrow$ (iii)), (11.5) is equivalent to (11.6). Conversely, if (11.6) holds, then for any nonsingular $A \in \mathbf{M}(D;I)$, $AA^t \in \mathbf{P}(D;I)$. By (11.5), $A^tA \in \mathbf{P}(D;I)^{-1}$, hence $A^{-1}(A^{-1})^t \in \mathbf{P}(D;I)$. Thus, again by (11.6), there exists $B \in \mathbf{M}(D;I)$ such that $A^{-1}(A^t)^{-1} = BB^t$, hence $AB(AB)^t = 1_I$, i.e., $\Gamma := AB$ is an orthogonal matrix. Since $\Gamma \in \overline{\mathbf{M}}(D;I) = \mathbf{M}(T(D);I)$ (Lemma 10.2) and $\mathbf{M}(T(D);I)$ is an algebra of block-triangular matrices (see Remark 10.3 and apply Remark 2.1 of [AP] (1993)), Γ must be block-diagonal, i.e., $u \neq v \Rightarrow \Gamma_{[uv]} = 0$ $(u, v \in V)$. Therefore $A^{-1} \equiv B\Gamma^{-1} \in \mathbf{M}(D;I)$; thus $\mathbf{M}(D;I)$ is closed under matrix inversion, so D is transitive by Lemma 10.1 ((iii) $\Rightarrow$ (i)).

12. The maximal normal linear ADG model determined by a multivariate regression subspace.

In a specific application with a multivariate observation space $\mathbf{M}(I \times N)$ and covariance structure of the form $\Sigma \otimes 1_N$, one might encounter a multivariate linear regression subspace $L \subseteq \mathbf{M}(I \times N)$ of the form

(12.1) $$L = \times(\, L_v \mid v \in V),$$

where V is an index set determining a partitioning

(12.2) $$I = \dot{\cup}(\, I_v \mid v \in V),\ I_v \neq \emptyset,$$

and where for each $v \in V$,

(12.3) $$L_v \text{ is a MANOVA subspace of } \mathbf{M}(I_v \times N).$$

In general, $(L_v | v \in V)$ may be a nonnested family of MANOVA subspaces, in the sense that $(K_{L_v} | v \in V)$ is a family of nonnested subspaces of $\mathbf{R}^N$ - recall (6.2). If the covariance matrix Σ is unrestricted, this constitutes Zellner's (1982) seemingly unrelated regressions (SUR) model, which does not admit explicit MLEs unless $(K_{L_v} | v \in V)$ is actually nested. We now show how to determine a parsimonious set of covariance restrictions of ADG Markov form such that the resulting multivariate normal linear regression model is a normal linear ADG model, hence admits explicit MLEs as in Section 7. More precisely, we show how to construct the unique maximal ADG $D \equiv D(L)$ such that L is a D-subspace. The resulting normal linear ADG model $\mathbf{N}_{I\times N}(L, D(L))$ is maximal in the sense that if L is also an E-subspace for another ADG E, then $\mathbf{N}_{I\times N}(L, E)$ is a submodel of $\mathbf{N}_{I\times N}(L, D(L))$. Examples of this construction are given in §13.2.

For each $v \in V$, set $K_v = K_{L_v}$ (recall (6.2)). Without loss of generality we may assume that the K_v are distinct[6] subspaces of $\mathbf{R}^N$. Define $D(L)$ to be the TADG with vertex set V and transitive binary relation $\prec_L$ defined as follows (compare to (iv) above):

$$\forall u, v \in V,\ u \prec_L v \iff K_u \subset K_v. \tag{12.4}$$

By (12.1)-(12.4), L is a $D(L)$-subspace of $\mathbf{M}(I \times N)$. In fact, $D(L)$ is the maximal ADG with this property:

Proposition 12.1. Let $E \equiv (W, S)$ be another ADG with associated partitioning $I = \dot\cup(I_w | w \in w)$ such that L is also an E-subspace of $\mathbf{M}(I \times N)$. Then there exists a surjective ADG homomorphism $\psi\colon E \to D(L)$ that satisfies (9.1), hence $\mathbf{N}_{I\times N}(L, E)$ is a submodel of $\mathbf{N}_{I\times N}(L, D(L))$.

Proof. Define

$$\mathcal{J} := \left\{ J \subseteq I \;\middle|\; J \neq \emptyset,\ L_J \text{ is a MANOVA subspace of } \mathbf{M}(J \times N) \right\}.$$

Clearly $I_v \in \mathcal{J}\ \forall v \in V$. Note that if $J \in \mathcal{J}$ and $J' \subset J$, then $J' \in \mathcal{J}$ and $K_{L_{J'}} = K_{L_J}$. Since K_v, $v \in V$, are distinct, each $J \in \mathcal{J}$ therefore satisfies $J \subseteq I_v$ for exactly one $v \in V$. Thus the subsets I_v, $v \in V$, are the maximal elements of $\mathcal{J}$.

Since L is an E-subspace, $L_{[w]}$ is a MANOVA subspace of $\mathbf{M}([w] \times N)$ for each $w \in W$ (recall our notation $[w] := I_w$), hence $[w] \in \mathcal{J}$, so $[w] \subseteq I_v$ for exactly one $v \in V$. Define the mapping $\psi : W \to V$ as follows: $\psi(w) = v$ iff $[w] \subseteq I_v$. Clearly ψ is surjective and satisfies (9.1) and $K_{L_{[w]}} = K_{\psi(w)}$, $w \in W$. Furthermore, if $w \prec_E w'$ then $K_{L_{[w]}} \subseteq K_{L_{[w']}}$ since L is an E-subspace (recall Remark 6.2(iii)$''$), hence $K_{\psi(w)} \subseteq K_{\psi(w')}$. It follows from (12.4) that $\psi(w) \preceq_L \psi(w')$, hence ψ is an ADG homomorphism. The final assertion follows from Proposition 3.1.

Remark 12.1. If $(K_v | v \in V)$ is a nested family of subspaces, i.e., totally ordered under inclusion, then the maximal ADG $D(L)$ constructed according to (12.4) is a *complete* graph - every pair of its nodes is linked by an edge. Thus, under the resulting normal model $\mathbf{N}_{I\times N}(L, D(L))$, Σ is unrestricted, i.e., $\mathbf{P}(D(L); I) = \mathbf{P}(I)$. On the other hand, if no pairwise inclusions hold in the family $(K_v | v \in V)$, then $D(L)$ has *no* edges, so Σ has a block-diagonal form and the normal random variates $x_{[v]}$, $v \in V$, are mutually independent under the model $\mathbf{N}_{I\times N}(L, D(L))$.

Remark 12.2. In [AP] (1994, Section 6) it was shown how to construct $\mathcal{K}(L)$, the minimal ring of subsets of I such that L is a $\mathcal{K}(L)$-subspace of $\mathbf{M}(I \times N)$. The resulting normal LCI model $\mathbf{N}_{I\times N}(L, \mathcal{K}(L))$ was shown there to be maximal in the sense that if L is also an $\mathcal{M}$-subspace for a ring $\mathcal{M}$ of subsets of I, then $\mathbf{N}_{I\times N}(L, \mathcal{M})$ is a submodel of $\mathbf{N}_{I\times N}(L, \mathcal{K}(L))$. By the 1-1 correspondence between LCI models and TADG models established by [AMPT]

[6] If, instead, $K_u = K_v$ for some $u \neq v$, simply replace u and v by a single new element called $u \cup v$ and redefine V accordingly, then define $I_{u\cup v} := I_u \cup I_v$ and $L_{u\cup v} := L_u \times L_v \subseteq \mathbf{M}(I_{u\cup v} \times N)$, again a MANOVA subspace. Then $K_{u\cup v} = K_u = K_v$.

(1997, Theorem 4.1), this result also follows from the stronger result in Proposition 12.1 - in fact, $\mathcal{K}(L) = \mathcal{K}(D(L))$ and $\mathbf{N}_{I\times N}(L, \mathcal{K}(L)) = \mathbf{N}_{I\times N}(L, D(L))$.

We now generalize this construction of the maximal ADG $D(L)$ as follows. Suppose that we have not one but two multivariate linear regression subspaces $L, M \subseteq \mathbf{M}(I\times N)$ of the form given by (12.1)-(12.3):

$$L = \times(L_v|v \in V)$$
$$M = \times(M_w|w \in W)\ ,$$

where V, W are index sets that determine two partitionings

$$I = \dot\cup(I_v|v \in V) = \dot\cup(I_w|w \in W), \quad I_v, I_w \neq \emptyset,$$

and where $(L_v|v \in V)$, $(M_w|w \in W)$ are two families of possibly nonnested MANOVA subspaces with

$$L_v \subseteq \mathbf{M}(I_v\times N)$$
$$M_w \subseteq \mathbf{M}(I_w\times N)\ .$$

Suppose that we wish to determine a parsimonious set of covariance restrictions of ADG Markov form such that both resulting multivariate normal linear regression models are normal linear ADG models. That is, we wish to construct the maximal ADG $D \equiv D(L, M)$ such that both L and M are D-subspaces. The resulting normal linear ADG models $\mathbf{N}_{I\times N}(L, D(L, M))$ and $\mathbf{N}_{I\times N}(M, D(L, M))$ are simultaneously maximal in the sense that if L and M are also E-subspaces for another ADG E, then $\mathbf{N}_{I\times N}(L, E)$ and $\mathbf{N}_{I\times N}(M, E)$ must be submodels of $\mathbf{N}_{I\times N}(L, D(L, M))$ and $\mathbf{N}_{I\times N}(M, D(L, M))$, respectively.

This situation typically occurs when $M \subset L$ and we wish to test $\xi \in M$ *vs.* $\xi \in L$ as in Section 9. However, the following construction of $D(L, M)$ does not require that $M \subset L$.

The ADG $D(L, M)$ is defined to have vertex set

$$V(L, M) := \big\{(v, w) \in V\times W \;\big|\; I_v \cap I_w \neq \emptyset\big\}, \tag{12.5}$$

with the associated partitioning of the index set I given by

$$I = \dot\cup\big(\, I_{(v,w)} \mid (v, w) \in V(L, M)\,\big), \tag{12.6}$$

where, for $(v, w) \in V(L, M)$,

$$I_{(v,w)} := I_v \cap I_w =: [v, w] \neq \emptyset.$$

Then

$$\begin{aligned} L &= \times\big(\, L_{[v,w]} \mid (v, w) \in V(L, M)\,\big),\\ M &= \times\big(\, M_{[v,w]} \mid (v, w) \in V(L, M)\,\big), \end{aligned} \tag{12.7}$$

and for each $(v,w)\in V(L,M)$, $L_{[v,w]}$ and $M_{[v,w]}$ are MANOVA subspaces of $\mathbf{M}([v,w]\times N)$ such that

$$K_{L_{[v,w]}} = K_{L_v}$$
$$K_{M_{[v,w]}} = K_{M_w}. \tag{12.8}$$

As before, we may assume that $(K_{L_v}|v\in V)$ and $(K_{M_w}|w\in W)$ are families of distinct subspaces of $\mathbf{R}^N$, so that $\big((K_{L_v},K_{M_w})\mid(v,w)\in V(L,M)\big)$ is a family of distinct pairs of subspaces of $\mathbf{R}^N$, i.e.,

$$(v,w)\neq(v',w')\Rightarrow K_{L_v}\neq K_{L_{v'}} \text{ or } K_{M_w}\neq K_{M_{w'}} \text{ (or both).} \tag{12.9}$$

Now define $D(L,M)$ to be the TADG with vertex set $V(L,M)$ and transitive binary relation $\prec_{L,M}$ defined as follows:
(12.10)

$$\forall(v,w),(v',w')\in V(L,M),\quad (v,w)\prec_{L,M}(v',w') \iff (K_{L_v},K_{M_w})\subset(K_{L_{v'}},K_{M_{w'}}),$$

where $(K_{L_v},K_{M_w})\subset(K_{L_{v'}},K_{M_{w'}})$ is defined to mean that $K_{L_v}\subseteq K_{L_{v'}}$ and $K_{M_w}\subseteq K_{M_{w'}}$ with at least one proper inclusion. It follows from (12.6)-(12.8) and (12.10) that both L and M are $D(L,M)$-subspaces of $\mathbf{M}(I\times N)$. We now show that $D(L,M)$ is the unique maximal ADG with this property.

Proposition 12.2. Let $E\equiv(U,S)$ be another ADG, with associated partitioning $I=\dot\cup(I_u|u\in U)$ such that L and M are also E-subspaces of $\mathbf{M}(I\times N)$. Then there exists a surjective ADG homomorphism $\psi\colon E\to D(L,M)$ such that

$$I_{(v,w)}=\dot\cup\big(I_u \,\big|\, u\in U,\psi(u)=(v,w)\big),\ v\in V, \tag{12.11}$$

hence $\mathbf{N}_{I\times N}(L,E)$ is a submodel of $\mathbf{N}_{I\times N}(L,D(L))$.

Proof. Define

$$\mathcal{J}:=\big\{\,J\subseteq I \,\big|\, J\neq\emptyset,\ L_J \text{ and } M_J \text{ are MANOVA subspaces of } \mathbf{M}(J\times N)\,\big\}.$$

Clearly $I_{(v,w)}\in\mathcal{J}\ \forall(v,w)\in V(L,M)$. Note that if $J\in\mathcal{J}$ and $J'\subset J$, then $J'\in\mathcal{J}$ and $K_{L_{J'}}=K_{L_J}$, $K_{M_{J'}}=K_{M_J}$. By (12.9), each $J\in\mathcal{J}$ therefore satisfies $J\subseteq I_{(v,w)}$ for exactly one $(v,w)\in V(L,M)$. Thus the subsets $I_{(v,w)}$, $(v,w)\in V(L,M)$, are the maximal elements of $\mathcal{J}$.

Since L and M are E-subspaces, $L_{[u]}$ and $M_{[u]}$ are MANOVA subspaces of $\mathbf{M}([u]\times N)$ for each $u\in U$ ($[u]:=I_u$). Therefore $[u]\in\mathcal{J}$ and $[u]\subseteq I_{(v,w)}$ for exactly one $(v,w)\equiv(v(u),w(u))\in V(L,M)$, so we may define $\psi:U\to V(L,M)$ by $\psi(u)=(v(u),w(u))$. Then ψ is surjective and satisfies (12.11) and, for $w\in W$,

$$K_{L_{[u]}} = K_{L_{v(u)}}$$
$$K_{M_{[u]}} = K_{M_{w(u)}}. \tag{12.12}$$

Furthermore, if $u \prec_E u'$ then

$$K_{L_{[u]}} \subseteq K_{L_{[u']}} \\ K_{M_{[u]}} \subseteq K_{M_{[u']}} \tag{12.13}$$

since L and M are E-subspaces (recall Remark 6.2(iii)$''$), so $\psi(u) \preceq_{L,M} \psi(u')$ by (12.10) and (12.12). Thus ψ is an ADG homomorphism. The final assertion again follows from Proposition 3.1.

This construction of $D(L, M)$ can be extended in an obvious way to the case of three or more regression subspaces of the form given by (12.1)-(12.3).

Remark 12.3. For a fixed ADG D, Definition 6.2, Remark 6.2, and Proposition 10.1 provide several characterizations of the set of all D-subspaces. Two ADGs D, D' with the same vertex set V are called *Markov equivalent* if they determine the same Markov model, i.e., if $\mathcal{P}(D; \mathbf{X}) = \mathcal{P}(D'; \mathbf{X})$ for all $\mathbf{X}$. A discussion of Markov equivalence of ADGs, including the necessary and sufficient graphical condition for Markov equivalence, can be found in [AMP] (1997b). If D and D' are Markov equivalent but not identical, the set of all D-subspaces will not be identical to the set of all D'-subspaces. For example, using notation similar to that of Example 4 in Section 13, $D := 1 \to 2$ and $D' := 1 \leftarrow 2$ both determine the same (vacuous) Markov condition, but $L \equiv L_{[1]} \times L_{[2]}$ is a D-subspace (resp., D'-subspace) iff the associated subspaces $K_1, K_2 \subseteq \mathbf{R}^N$ satisfy $K_1 \subseteq K_2$ (resp., $K_1 \supseteq K_2$). Similarly, $D := 1 \to 2 \to 3$ and $D' := 1 \leftarrow 2 \leftarrow 3$ both determine the same (non-trivial) Markov condition $1 \perp\!\!\!\perp 3 \mid 2$, but $L \equiv L_{[1]} \times L_{[2]} \times L_{[3]}$ is a D-subspace (resp., D'-subspace) iff the associated subspaces $K_1, K_2, K_3 \subseteq \mathbf{R}^N$ satisfy $K_1 \subseteq K_2 \subseteq K_3$ (resp., $K_1 \supseteq K_2 \supseteq K_3$).

Let $[D]$ denote the Markov-equivalence class of ADGs that contains D. Note that if $D' \in [D]$ then $\mathbf{N}_I(D) = \mathbf{N}_I(D')$, hence we denote this normal covariance model by $\mathbf{N}_I([D])$. We say that L is a $[D]$-*subspace* if L is a D'-subspace for some $D' \in [D]$; in this case $\mathbf{N}_{I \times N}(L, D')$ is a normal linear ADG model with the same covariance structure as $\mathbf{N}_I([D])$. It is therefore of statistical interest to characterize the set of all $[D]$-subspaces. A characterization that depends on the *essential graph* D^* associated with $[D]$ (cf. [AMP] (1997b)) will be presented in a subsequent paper.

13. Examples.

In §13.1, nine examples of normal linear ADG models $\mathbf{N}_{I \times N}(L, D)$ as defined and developed in Sections 4, 6, and 7 are presented. In each example the index sets I and N are $\{1, 2, 3, 4\}$ and $\{1, \dots, n\}$ ($n \geq 4$), respectively, so that the model consists of n independent 4-variate normal observations $x_1, \dots, x_n$, each with the same unknown covariance matrix $\Sigma \in \mathbf{P}(D; I)$; as before, set

$$y := (x_1, \dots, x_n) \in \mathbf{M}(I \times N)$$

and $\xi := \mathrm{E}(y)$. For each model $\mathbf{N}_{I \times N}(L, D)$, we specify the ADG $D \equiv (V, R)$, the associated partitioning $I = \dot\cup(I_v | v \in V)$, and the Markov conditions they determine;

the D-linear regression subspace $L \subseteq \mathbf{M}(I \times N)$, its representation $L = \times(L_{[v]}|v \in V)$ as a product of MANOVA subspaces, the linear subspace $K_v \subseteq \mathbf{R}^N$ associated with each $L_{[v]}$ (cf. (6.2)), and $p_v \equiv \dim(K_v)$; the D-parameters $\overline{\pi}_D(\xi, \Sigma) \equiv ((\mu_v, \beta_v, \Lambda_v)|v \in V)$ and the necessary and sufficient condition (7.2) for the a.e. existence of the MLE $(\hat{\xi}, \hat{\Sigma})$ under the model. In Examples 4-9 we also apply the algebraic characterization in Proposition 10.1 to verify that L is a D-subspace.

In §13.2, we illustrate the construction (12.4) of the maximal ADG $D(L)$ determined by a linear subspace $L \subseteq \mathbf{M}(I \times N)$ given by (12.1)-(12.3).

In §13.3, nested pairs of the models in Examples 2-9 are selected to illustrate the general testing problem (9.2) for normal linear ADG models in Section 9. In Example 10 we show that in a 4-variate two-way MANOVA model with no interactions, under a suitable ADG Markov assumption it is possible simultaneously to test the hypotheses of no row effects for one variable, no column effects for a second, and no row or column effects for a third.

13.1. Normal linear ADG models.

In Examples 1-9, the D-subspace L is taken to be of the form

$$L \equiv L(\mathbf{B}) := \{\, BZ \mid B \in \mathbf{B} \,\} \tag{13.1}$$

(compare to (6.4)), where

$$B \equiv \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{pmatrix} \in \mathbf{M}(I \times T)$$

is a matrix of unknown regression coefficients ($T := \{1, 2, 3, 4\}$),

$$Z \equiv \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ z_{31} & z_{32} & \cdots & z_{3n} \\ z_{41} & z_{42} & \cdots & z_{4n} \end{pmatrix} \equiv \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} \in \mathbf{M}(T \times N)$$

is a known design matrix of full rank 4, and $\mathbf{B}$ is a linear subspace of $\mathbf{M}(I \times T)$ that determines L.

Example 1. Let $D \equiv (V, R)$ be the trivial ADG with only one node, i.e., $V := \{v_{1234}\}$, and $R = \emptyset$. For convenience we abbreviate v_{1234} by 1234. Then I is partitioned trivially as

$$I = I_{1234} \equiv [1234] := \{1, 2, 3, 4\} \tag{13.2}$$

and the Markov condition is vacuous, so $\mathbf{P}(D; I) = \mathbf{P}(I)$, i.e., Σ is unrestricted. Set $\mathbf{B} = \mathbf{M}(I \times T)$ in (13.1), i.e., B is unrestricted, so that

$$L = \mathbf{R}^I \otimes K \tag{13.3}$$

where

$$K = \mathrm{row}(Z) \equiv \mathrm{span}(z_1, z_2, z_3, z_4). \tag{13.4}$$

Then $L_{[1234]} = L$, a MANOVA subspace, and the corresponding subspace $K_{1234} = K$. Thus, trivially, L is a D-subspace by Definition 6.2 and $\mathbf{N}_{I\times N}(L, D)$ is a normal linear ADG model, namely the MANOVA model $\mathbf{N}_{I\times N}(L)$. Since $\prec 1234 \succ = \emptyset$, for $(\xi, \Sigma) \in L \times \mathbf{P}(I)$ the D-parameters $\overline{\pi}(\xi, \Sigma)$ are

$$\mu_{1234} = \xi, \quad \beta_{1234} = 0, \quad \Lambda_{1234} = \Sigma. \tag{13.5}$$

Since $p_{1234} := \mathrm{tr}(P_{1234}) = 4$ and $| \preceq 1234 \succeq | = 4$, condition (7.2) for the existence of the MLEs $\hat{\xi}, \hat{\Sigma}$ is $n \geq 8$.

Example 2. Here we consider the same model $\mathbf{N}_{I\times N}(L, D)$ as in Example 1 but with a different (finer) parameterization. Let $V := \{v_1, v_{23}, v_4\} \equiv \{1, 23, 4\}$ and let D be the following ADG:

$$\begin{array}{ccc} & 4 & \\ \nearrow & & \nwarrow \\ 1 & \longrightarrow & 23 \end{array} .$$

Partition I as

$$I = I_1 \dot{\cup} I_{23} \dot{\cup} I_4 \equiv [1] \dot{\cup} [23] \dot{\cup} [4] := \{1\} \dot{\cup} \{2, 3\} \dot{\cup} \{4\}. \tag{13.6}$$

Again the Markov condition is vacuous, so $\mathbf{P}(D; I) = \mathbf{P}(I)$. Take $\mathbf{B}, L, K$ as in Example 1, but represent L equivalently as

$$L = L_{[1]} \times L_{[23]} \times L_{[4]}, \tag{13.7}$$

where

$$\begin{aligned} L_{[1]} &= \mathbf{R}^{[1]} \otimes K_1 \\ L_{[23]} &= \mathbf{R}^{[23]} \otimes K_{23} \\ L_{[4]} &= \mathbf{R}^{[4]} \otimes K_4 \ , \end{aligned} \tag{13.8}$$

with

$$K_1 = K_{23} = K_4 = K \equiv \mathrm{span}\,(z_1, z_2, z_3, z_4) \ . \tag{13.9}$$

That L is a D-subspace is immediate from Definition 6.2 and Remark 6.2(iii)$''$. Since $\prec 1 \succ = \emptyset$, $\prec 23 \succ = \{1\}$, and $\prec 4 \succ = \{1, 2, 3\}$, for $(\xi, \Sigma) \in L \times \mathbf{P}(I)$ the D-parameters $\overline{\pi}(\xi, \Sigma)$ are

(13.10)

$$\begin{array}{lll}
\mu_1 = \xi_{\{1\}} & \beta_1 = 0 & \Lambda_1 = \Sigma_{\{1\}} \\
\mu_{23} = \xi_{\{2,3\}} - \Sigma_{\{2,3\}\{1\}}\Sigma^{-1}_{\{1\}}\xi_{\{1\}} & \beta_{23} = \Sigma_{\{2,3\}\{1\}}\Sigma^{-1}_{\{1\}} & \Lambda_{23} = \Sigma_{\{2,3\}\bullet\{1\}} \\
\mu_4 = \xi_{\{4\}} - \Sigma_{\{4\}\{1,2,3\}}\Sigma^{-1}_{\{1,2,3\}}\xi_{\{1,2,3\}} & \beta_4 = \Sigma_{\{4\}\{1,2,3\}}\Sigma^{-1}_{\{1,2,3\}} & \Lambda_4 = \Sigma_{\{4\}\bullet\{1,2,3\}}\ .
\end{array}$$

Since $p_1 = p_{23} = p_4 = 4$ and $|\preceq 1 \succeq| = 1$, $|\preceq 23 \succeq| = 2$, $|\preceq 4 \succeq| = 4$, condition (7.2) again becomes $n \geq 8$.

The MANOVA model in Examples 1 and 2 properly contains each of the normal linear ADG models in Examples 3-9.

Example 3. Let the ADG D and the associated partitioning of I be as in Example 2, so again $\mathbf{P}(D; I) = \mathbf{P}(I)$. Now define

$$\mathbf{B} = \{B \mid b_{12} = b_{13} = b_{14} = b_{24} = b_{34} = 0\},$$

so that $\mathbf{B}$ is the set of all $B \in \mathbf{M}(I \times T)$ of the form

$$B \equiv \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ b_{21} & b_{22} & b_{23} & 0 \\ b_{31} & b_{32} & b_{33} & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} \end{pmatrix}.$$

Again L is given by (13.7) and (13.8), where now

$$\begin{aligned} K_1 &= \text{span}(z_1) \\ K_{23} &= \text{span}(z_1, z_2, z_3) \\ K_4 &= \text{span}(z_1, z_2, z_3, z_4)\ . \end{aligned} \tag{13.11}$$

Then

$$K_1 \subset K_{23} \subset K_4, \tag{13.12}$$

so it follows from Definition 6.2 and Remark 6.2(iii)$''$ that L is a D-subspace. (Conversely, were $K_1 \not\subseteq K_{23}$ or $K_{23} \not\subseteq K_4$, L would not be a D-subspace, since (iii)$''$ would be violated.) Because D is the same as in Example 2, for $(\xi, \Sigma) \in L \times \mathbf{P}(I)$ the D-parameters $\overline{\pi}(\xi, \Sigma)$ remain the same as in (13.10) (although the ranges $L_{[1]}$ and $L_{[23]}$ of μ_1 and μ_{23} are different). Since $p_1 = 1$, $p_{23} = 3$, $p_4 = 4$ and again $|\preceq 1 \succeq| = 1$, $|\preceq 23 \succeq| = 2$, $|\preceq 4 \succeq| = 4$, condition (7.2) remains $n \geq 8$.

In Examples 4-9 we take $V := \{v_1, v_2, v_3, v_4\} \equiv \{1, 2, 3, 4\}$ and partition I simply as

$$I = I_1 \dot\cup I_2 \dot\cup I_3 \dot\cup I_4 \equiv [1] \dot\cup [2] \dot\cup [3] \dot\cup [4] := \{1\} \dot\cup \{2\} \dot\cup \{3\} \dot\cup \{4\}. \tag{13.13}$$

Example 4. Let D be the following ADG:

$$\begin{array}{ccccc} 1 & \longrightarrow & 3 & \longleftarrow & 2 \\ & & \downarrow & \swarrow & \\ & & 4 & & \end{array} .$$

From Definition 4.2, the Markov conditions determined by D and the above partitioning of I are $1 \perp\!\!\!\perp 2$ and $4 \perp\!\!\!\perp 1 \mid 2, 3$, which determine $\mathbf{P}(D; I)$ according to Definition 4.2. Define

$$\mathbf{B} = \{B \mid b_{12} = b_{13} = b_{14} = b_{21} = b_{23} = b_{24} = b_{34} = 0\},$$

so that $\mathbf{B}$ is the set of all $B \in \mathbf{M}(I \times T)$ of the form

$$B \equiv \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ b_{31} & b_{32} & b_{33} & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} \end{pmatrix} .$$

Now

$$L = L_{[1]} \times L_{[2]} \times L_{[3]} \times L_{[4]}, \tag{13.14}$$

where

$$L_{[i]} = \mathbf{R}^{[i]} \otimes K_i, \;\; i = 1, 2, 3, 4, \tag{13.15}$$

with

$$\begin{aligned} K_1 &= \text{span}(z_1) \\ K_2 &= \text{span}(z_2) \\ K_3 &= \text{span}(z_1, z_2, z_3) \\ K_4 &= \text{span}(z_1, z_2, z_3, z_4) \; . \end{aligned} \tag{13.16}$$

Then

$$\begin{aligned} K_1 \subset K_3 \subset K_4 \\ K_2 \subset K_3 \subset K_4 \; , \end{aligned} \tag{13.17}$$

so it follows from Definition 6.2 and Remark 6.2(iii)$''$ that L is a D-subspace. Since $\prec 1 \succ = \prec 2 \succ = \emptyset$, $\prec 3 \succ = \{1, 2\}$, and $\prec 4 \succ = \{2, 3\}$, for $(\xi, \Sigma) \in L \times \mathbf{P}(D; I)$ the D-parameters $\overline{\pi}(\xi, \Sigma)$ are

$$\begin{array}{lll} \mu_1 = \xi_{\{1\}} & \beta_1 = 0 & \Lambda_1 = \Sigma_{\{1\}} \\ \mu_2 = \xi_{\{2\}} & \beta_2 = 0 & \Lambda_2 = \Sigma_{\{2\}} \\ \mu_3 = \xi_{\{3\}} - \Sigma_{\{3\}\{1,2\}} \Sigma^{-1}_{\{1,2\}} \xi_{\{1,2\}} & \beta_3 = \Sigma_{\{3\}\{1,2\}} \Sigma^{-1}_{\{1,2\}} & \Lambda_3 = \Sigma_{\{3\} \bullet \{1,2\}} \\ \mu_4 = \xi_{\{4\}} - \Sigma_{\{4\}\{2,3\}} \Sigma^{-1}_{\{2,3\}} \xi_{\{2,3\}} & \beta_4 = \Sigma_{\{4\}\{2,3\}} \Sigma^{-1}_{\{2,3\}} & \Lambda_4 = \Sigma_{\{4\} \bullet \{2,3\}} \; . \end{array} \tag{13.18}$$

Since $p_1 = p_2 = 1$, $p_3 = 3$, $p_4 = 4$ and $|\preceq 1 \succeq| = |\preceq 2 \succeq| = 1$, $|\preceq 3 \succeq| = |\preceq 4 \succeq| = 3$, condition (7.2) becomes $n \geq 7$.

To illustrate the use of Proposition 10.1 to verify that L is a D-subspace, first note that the linear class $\mathbf{M}(D;I)$ defined in (10.1) here consists of all $A \in \mathbf{M}(I)$ of the form

$$A = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ 0 & a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

Then $\mathbf{M}(D;I)\mathbf{B} \subseteq \mathbf{B}$, so $\mathbf{M}(D;I)L \subseteq L$ (recall (13.1)), hence L is a D-subspace by (10.4).

Example 5. Let D be the following ADG:

$$\begin{array}{ccc} 1 \longrightarrow & 3 & \longleftarrow 2 \\ & \downarrow & \\ & 4 & \end{array} .$$

The Markov conditions determined by D are $1 \perp\!\!\!\perp 2$ and $4 \perp\!\!\!\perp 1, 2 \mid 3$. Define $\mathbf{B}$ and L as in Example 4, so that $L_{[i]}$, K_i, and p_i, $i = 1, 2, 3, 4$, remain the same as in that example. By (13.17), it follows from Definition 6.2 and Remark 6.2(iii)$''$ that L is a D-subspace. Again $\prec 1 \succ = \prec 2 \succ = \emptyset$ and $\prec 3 \succ = \{1, 2\}$ but now $\prec 4 \succ = \{3\}$, so the D-parameters $(\,(\mu_i, \beta_i, \Lambda_i) \mid i = 1, 2, 3)$ remain the same as in Example 2 but now

$$\mu_4 = \xi_{\{4\}} - \Sigma_{\{4\}\{3\}}\Sigma_{\{3\}}^{-1}\xi_{\{3\}}, \quad \beta_4 = \Sigma_{\{4\}\{3\}}\Sigma_{\{3\}}^{-1}, \quad \Lambda_4 = \Sigma_{\{4\}\bullet\{3\}}. \tag{13.19}$$

Also $|\preceq 1 \succeq| = |\preceq 2 \succeq| = 1$ and $|\preceq 3 \succeq| = 3$, while now $|\preceq 4 \succeq| = 2$, so condition (7.2) changes to $n \geq 6$.

The transitive closures $T(D)$ are identical for the ADGs D in Examples 4 and 5, i.e., in both examples $T(D)$ is the TADG

$$\begin{array}{ccc} 1 \longrightarrow & 3 & \longleftarrow 2 \\ \searrow & \downarrow & \swarrow \\ & 4 & \end{array} . \tag{13.20}$$

It follows that the families of D-subspaces coincide in Examples 4 and 5.

Example 6. Let D be the following ADG:

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

The Markov conditions determined by D are $3 \perp\!\!\!\perp 1 \mid 2$ and $4 \perp\!\!\!\perp 2 \mid 1, 3$. Define

$$\mathbf{B} = \{B \mid b_{11} = b_{12} = b_{13} = b_{14},\ b_{22} = b_{23} = b_{24},\ b_{33} = b_{34}\},$$

so that $\mathbf{B}$ is the set of all $B \in \mathbf{M}(I \times T)$ of the form

$$B \equiv \begin{pmatrix} a_1 & a_1 & a_1 & a_1 \\ b_1 & b_2 & b_2 & b_2 \\ c_1 & c_2 & c_3 & c_3 \\ d_1 & d_2 & d_3 & d_4 \end{pmatrix}.$$

Then L is given by (13.14) and (13.15) where now

$$\begin{aligned} K_1 &= \operatorname{span}(z_1 + z_2 + z_3 + z_4) \\ K_2 &= \operatorname{span}(z_1, z_2 + z_3 + z_4) \\ K_3 &= \operatorname{span}(z_1, z_2, z_3 + z_4) \\ K_4 &= \operatorname{span}(z_1, z_2, z_3, z_4)\ . \end{aligned} \tag{13.21}$$

Since

$$K_1 \subset K_2 \subset K_3 \subset K_4, \tag{13.22}$$

it follows from Definition 6.2 and Remark 6.2(iii)$''$ that L is a D-subspace. Since $\prec 1 \succ = \emptyset$, $\prec 2 \succ = \{1\}$, $\prec 3 \succ = \{2\}$, and $\prec 4 \succ = \{1,3\}$, for $(\xi, \Sigma) \in L \times \mathbf{P}(D; I)$ the D-parameters $\overline{\pi}(\xi, \Sigma)$ are

$$\begin{aligned} \mu_1 &= \xi_{\{1\}} & \beta_1 &= 0 & \Lambda_1 &= \Sigma_{\{1\}} \\ \mu_2 &= \xi_{\{2\}} - \Sigma_{\{2\}\{1\}}\Sigma_{\{1\}}^{-1}\xi_{\{1\}} & \beta_2 &= \Sigma_{\{2\}\{1\}}\Sigma_{\{1\}}^{-1} & \Lambda_2 &= \Sigma_{\{2\}\bullet\{1\}} \\ \mu_3 &= \xi_{\{3\}} - \Sigma_{\{3\}\{2\}}\Sigma_{\{2\}}^{-1}\xi_{\{2\}} & \beta_3 &= \Sigma_{\{3\}\{2\}}\Sigma_{\{2\}}^{-1} & \Lambda_3 &= \Sigma_{\{3\}\bullet\{2\}} \\ \mu_4 &= \xi_{\{4\}} - \Sigma_{\{4\}\{1,3\}}\Sigma_{\{1,3\}}^{-1}\xi_{\{1,3\}} & \beta_4 &= \Sigma_{\{4\}\{1,3\}}\Sigma_{\{1,3\}}^{-1} & \Lambda_4 &= \Sigma_{\{4\}\bullet\{1,3\}}\ . \end{aligned} \tag{13.23}$$

Since $p_i = i, i = 1,2,3,4$ and $|\preceq 1 \succeq| = 1$, $|\preceq 2 \succeq| = |\preceq 3 \succeq| = 2$, $|\preceq 4 \succeq| = 3$, (7.2) becomes $n \geq 7$.

Here the linear class $\mathbf{M}(D; I)$ in (10.1) consists of all $A \in \mathbf{M}(I)$ of the form

$$A = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & a_{32} & a_{33} & 0 \\ a_{41} & 0 & a_{43} & a_{44} \end{pmatrix}.$$

Then $\mathbf{M}(D; I)\mathbf{B} \subseteq \mathbf{B}$, so $\mathbf{M}(D; I)L \subseteq L$ (recall (13.1)), hence L is a D-subspace by Proposition 10.1.

Example 7. Let D be the following ADG:

$$\begin{matrix} 1 & & 4 \\ \downarrow & & \uparrow \\ 2 & \longrightarrow & 3 \end{matrix}\ .$$

The Markov conditions determined by D are $3 \perp\!\!\!\perp 1 \mid 2$ and $4 \perp\!\!\!\perp 1,2 \mid 3$. Let $\mathbf{B}$, L, and $L_{[i]}, K_{[i]}$, $i = 1,2,3,4$, be the same as in Example 6, so again L is a D-subspace. Since $\prec 1 \succ = \emptyset$, $\prec 2 \succ = \{1\}$, $\prec 3 \succ = \{2\}$ but now $\prec 4 \succ = \{3\}$, the D-parameters $(\,(\mu_i, \beta_i, \Lambda_i) \mid i = 1,2,3)$ remain the same as in Example 6 but now

$$\mu_4 = \xi_{\{4\}} - \Sigma_{\{4\}\{3\}}\Sigma^{-1}_{\{3\}}\xi_{\{3\}}, \quad \beta_4 = \Sigma_{\{4\}\{3\}}\Sigma^{-1}_{\{3\}}, \quad \Lambda_4 = \Sigma_{\{4\}\bullet\{3\}} \; . \tag{13.24}$$

Since $p_i = i, i = 1,2,3,4$ and $|\preceq 1 \succeq| = 1$, $|\preceq 2 \succeq| = |\preceq 3 \succeq| = |\preceq 4 \succeq| = 2$, (7.2) becomes $n \geq 6$.

The transitive closures $T(D)$ are identical for the ADGs D in Examples 6 and 7, i.e., $T(D)$ is the TADG

$$\begin{array}{ccc} 1 & \longrightarrow & 4 \\ \downarrow & \times & \uparrow \\ 2 & \longrightarrow & 3 \end{array} \tag{13.25}$$

in both examples. Therefore, the families of D-subspaces coincide in Examples 6 and 7.

Example 8. Let D be the following ADG:

$$\begin{array}{ccc} 1 & \longrightarrow & 3 \\ \downarrow & & \downarrow \\ 2 & \longrightarrow & 4 \end{array} .$$

The Markov conditions determined by D are $2 \perp\!\!\!\perp 3 \mid 1$ and $4 \perp\!\!\!\perp 1 \mid 2,3$. Define

$$\mathbf{B} = \{B \mid b_{11} = b_{12} = b_{13} = b_{14},\; b_{21} = b_{22},\; b_{23} = b_{24},\; b_{31} = b_{33},\; b_{32} = b_{34}\},$$

so that $\mathbf{B}$ is the set of all $B \in \mathbf{M}(I \times T)$ of the form

$$B \equiv \begin{pmatrix} a_1 & a_1 & a_1 & a_1 \\ b_1 & b_1 & b_2 & b_2 \\ c_1 & c_2 & c_1 & c_2 \\ d_1 & d_2 & d_3 & d_4 \end{pmatrix} .$$

Then L can be expressed by (13.14) and (13.15), where now

$$\begin{aligned} K_1 &= \operatorname{span}(z_1 + z_2 + z_3 + z_4) \\ K_2 &= \operatorname{span}(z_1 + z_2, z_3 + z_4) \\ K_3 &= \operatorname{span}(z_1 + z_3, z_2 + z_4) \\ K_4 &= \operatorname{span}(z_1, z_2, z_3, z_4) \; . \end{aligned} \tag{13.26}$$

Since

$$\begin{aligned} &K_1 \subset K_2 \subset K_4 \\ &K_1 \subset K_3 \subset K_4 \; , \end{aligned} \tag{13.27}$$

it follows from Definition 6.2 and Remark 6.2(iii)$''$ that L is a D-subspace. Again $\prec 1 \succ = \emptyset$, $\prec 2 \succ = \{1\}$ but now $\prec 3 \succ = \{1\}$, $\prec 4 \succ = \{2,3\}$, the D-parameters $((\mu_i, \beta_i, \Lambda_i) \mid i = 1,2)$ remain the same as in Examples 6 and 7 but now

$$\begin{aligned} &\mu_3 = \xi_{\{3\}} - \Sigma_{\{3\}\{1\}}\Sigma^{-1}_{\{1\}}\xi_{\{1\}} && \beta_3 = \Sigma_{\{3\}\{1\}}\Sigma^{-1}_{\{1\}} && \Lambda_3 = \Sigma_{\{3\}\bullet\{1\}} \\ &\mu_4 = \xi_{\{4\}} - \Sigma_{\{4\}\{2,3\}}\Sigma^{-1}_{\{2,3\}}\xi_{\{2,3\}} && \beta_4 = \Sigma_{\{4\}\{2,3\}}\Sigma^{-1}_{\{2,3\}} && \Lambda_4 = \Sigma_{\{4\}\bullet\{2,3\}} \ . \end{aligned} \tag{13.28}$$

Since $p_1 = 1$, $p_2 = p_3 = 2$, $p_4 = 4$ and $|\preceq 1 \succeq| = 1$, $|\preceq 2 \succeq| = |\preceq 3 \succeq| = 2$, $|\preceq 4 \succeq| = 3$, (7.2) is $n \geq 7$.

In this example, the linear class $\mathbf{M}(D; I)$ in (10.1) consists of all $A \in \mathbf{M}(I)$ of the form

$$A = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 \\ 0 & a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

Here again $\mathbf{M}(D; I)\mathbf{B} \subseteq \mathbf{B}$, so $\mathbf{M}(D; I)L \subseteq L$ by (13.1) and Proposition 10.1 implies that L is a D-subspace.

Example 9. Let D be the following ADG:

$$\begin{array}{ccc} 1 & \longrightarrow & 3 \\ \downarrow & & \downarrow \\ 2 & & 4 \end{array} .$$

The Markov conditions determined by D are $2 \perp\!\!\!\perp 3,4 \mid 1$ and $4 \perp\!\!\!\perp 1,2 \mid 3$. Define

$$\mathbf{B} = \{B \mid b_{11} = b_{12} = b_{13} = b_{14},\ b_{21} = b_{22},\ b_{23} = b_{24},\ b_{31} = b_{33},\ b_{32} = b_{34},\ b_{41} = b_{43}\},$$

so that $\mathbf{B}$ is the set of all $B \in \mathbf{M}(I \times T)$ of the form

$$B \equiv \begin{pmatrix} a_1 & a_1 & a_1 & a_1 \\ b_1 & b_1 & b_2 & b_2 \\ c_1 & c_2 & c_1 & c_2 \\ d_1 & d_2 & d_1 & d_3 \end{pmatrix}.$$

Again L is given by (13.14) and (13.15), where now

$$\begin{aligned} K_1 &= \text{span}(z_1 + z_2 + z_3 + z_4) \\ K_2 &= \text{span}(z_1 + z_2, z_3 + z_4) \\ K_3 &= \text{span}(z_1 + z_3, z_2 + z_4) \\ K_4 &= \text{span}(z_1 + z_3, z_2, z_4) \ . \end{aligned} \tag{13.29}$$

Because

$$\begin{gathered} K_1 \subset K_2 \\ K_1 \subset K_3 \subset K_4 \ , \end{gathered} \tag{13.30}$$

it follows from Definition 6.2 and Remark 6.2(iii)$''$ that L is a D-subspace. Since $\prec 1 \succ = \emptyset$, $\prec 2 \succ = \{1\}$, $\prec 3 \succ = \{1\}$ but $\prec 4 \succ = \{3\}$, the D-parameters $(\,(\mu_i, \beta_i, \Lambda_i) \mid i = 1,2,3)$ remain the same as in Example 8 but here

$$\mu_4 = \xi_{\{4\}} - \Sigma_{\{4\}\{3\}}\Sigma_{\{3\}}^{-1}\xi_{\{3\}}, \quad \beta_4 = \Sigma_{\{4\}\{3\}}\Sigma_{\{3\}}^{-1}, \quad \Lambda_4 = \Sigma_{\{4\}\bullet\{3\}}\ . \tag{13.31}$$

Since $p_1 = 1$, $p_2 = p_3 = 2$, $p_4 = 3$ and $|\preceq 1 \succeq| = 1$, $|\preceq 2 \succeq| = |\preceq 3 \succeq| = |\preceq 4 \succeq| = 2$, (7.2) is $n \geq 5$.

In this example, the linear class $\mathbf{M}(D; I)$ in (10.1) consists of all $A \in \mathbf{M}(I)$ of the form

$$A = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 \\ 0 & 0 & a_{43} & a_{44} \end{pmatrix}.$$

Again $\mathbf{M}(D; I)\mathbf{B} \subseteq \mathbf{B}$ so $\mathbf{M}(D; I)L \subseteq L$, hence L is a D-subspace by Proposition 10.1.

13.2. Construction of the maximal ADG model $D(L)$.

In order to illustrate the construction of the maximal ADG $D(L)$ associated with a given regression subspace $L \subseteq \mathbf{M}(I \times N)$, first take L to be the subspace occurring in Examples 4 and 5. There we saw that L can be expressed in the form given by (13.14)-(13.16). By (12.4) and (13.17), $D(L)$ is the TADG in (13.20) with associated Markov condition $1 \perp\!\!\!\perp 2$, which is less restrictive than the Markov conditions determined by the ADGs D in Examples 4 and 5. Thus the normal linear ADG model $\mathbf{N}_{I\times N}(L, D(L))$ is less restrictive than the models $\mathbf{N}_{I\times N}(L, D)$ in these two examples, as guaranteed by Proposition 12.1.

Next take L to be the subspace occurring in Examples 6 and 7, so that L is given by (13.14), (13.15), and (13.21). By (12.4) and (13.22), $D(L)$ is the TADG in (13.25), a complete graph with vacuous Markov condition - recall Remark 12.1. Again, the normal linear ADG model $\mathbf{N}_{I\times N}(L, D(L))$ is less restrictive than the models $\mathbf{N}_{I\times N}(L, D)$ in Examples 6 and 7, as implied by Proposition 12.1.

Now take L be the subspace occurring in Example 8, so that now L is given by (13.14), (13.15), and (13.26). By (12.4) and (13.27), $D(L)$ is the TADG

$$\begin{array}{ccc} 1 & \longrightarrow & 3 \\ \downarrow & \searrow & \downarrow \\ 2 & \longrightarrow & 4 \end{array}\ . \tag{13.32}$$

with Markov condition $2 \perp\!\!\!\perp 3 \mid 1$, again less restrictive than the Markov condition determined by the ADG D in Example 8.

Lastly, let L be the subspace occurring in Example 9, so that L is given by (13.14), (13.15), and (13.29). Then by (12.4) and (13.30), $D(L)$ is the TADG

$$\begin{array}{ccc} 1 & \longrightarrow & 3 \\ \downarrow & \searrow & \downarrow \\ 2 & & 4 \end{array}\ . \tag{13.33}$$

with Markov conditions $2 \perp\!\!\!\perp 3,4 \mid 1$ and $4 \perp\!\!\!\perp 2 \mid 1,3$, again less restrictive than the Markov conditions determined by the ADG D in Example 9.

13.3. Testing normal linear ADG models.

We now exhibit nested pairs of the models in Examples 1-9 in order to illustrate the submodel relation in the general testing problem (9.2). Denote the ADG $D \equiv (V, R)$ and the D-subspace L occurring in Example i, $i = 1, \ldots, 9$ by $D_i \equiv (V_i, R_i)$ and L_i, respectively, and let $\mathbf{N}_i$ denote the normal linear ADG model $\mathbf{N}_{I\times N}(L_i, D_i)$. We shall verify the following relations among these models:

$$\begin{aligned} &\mathbf{N}_5 \subset \mathbf{N}_4 \subset \mathbf{N}_3 \subset \mathbf{N}_2 = \mathbf{N}_1 \\ &\mathbf{N}_7 \subset \mathbf{N}_6 \subset \mathbf{N}_2 \\ &\mathbf{N}_9 \subset \mathbf{N}_8 \subset \mathbf{N}_2\ . \end{aligned} \tag{13.34}$$

From their definitions,

$$\begin{aligned} L_5 = L_4 \subset L_3 \subset L_2 = L_1 \\ L_7 = L_6 \subset L_2 \\ L_9 \subset L_8 \subset L_2\ . \end{aligned}$$

Since $D_3 = D_2$, we have $\mathbf{N}_3 \subset \mathbf{N}_2$. To show that $\mathbf{N}(D_i) \subset \mathbf{N}(D_j)$ and therefore that $\mathbf{N}_i \subset \mathbf{N}_j$ in the remaining cases in (13.34), we must exhibit a proper surjective ADG homomorphism $\psi \equiv \psi_{ij} \colon D_i \to D_j$ that satisfies (9.1) with $W = V_i$, $V = V_j$.

For $(i, j) = (2, 1)$ we have $V_2 = \{1, 23, 4\}$ and $V_1 = \{1234\}$; thus we may define $\psi(1) = \psi(23) = \psi(4) := 1234$. Trivially, $\psi \colon D_2 \to D_1$ is a surjective ADG homomorphism that satisfies (9.1), hence $\mathbf{N}(D_2) \subseteq \mathbf{N}(D_1)$, but ψ is not proper and in fact $\mathbf{N}(D_2) = \mathbf{N}(D_1)$, so $\mathbf{N}_2 = \mathbf{N}_1$. For $(i, j) = (4, 3)$, $(6, 2)$, and $(8, 2)$ we have $V_i = \{1, 2, 3, 4\}$ and $V_j = \{1, 23, 4\}$; thus define $\psi(1) := 1$, $\psi(2) = \psi(3) := 23$, $\psi(4) := 4$. In each of these three cases it is readily verified that $\psi \colon D_i \to D_j$ is a proper surjective ADG homomorphism satisfying (9.1), hence $\mathbf{N}(D_i) \subset \mathbf{N}(D_j)$, so $\mathbf{N}_i \subset \mathbf{N}_j$ for $(i, j) = (4, 3)$, $(6, 2)$, $(8, 2)$. For $(i, j) = (5, 4)$, $(7, 6)$, and $(9, 8)$ we have $V_i = V_j = \{1, 2, 3, 4\}$; thus define $\psi := \mathrm{id}_{\{1,2,3,4\}}$. Again, in each case $\psi \colon D_i \to D_j$ is a proper surjective ADG homomorphism satisfying (9.1), hence $\mathbf{N}(D_i) \subset \mathbf{N}(D_j)$, so $\mathbf{N}_i \subset \mathbf{N}_j$ for $(i, j) = (5, 4)$, $(7, 6)$, $(9, 8)$.

For each nested pair $\mathbf{N}_i \subset \mathbf{N}_j$, the LR statistic λ_{ij} for testing $\mathbf{N}_i$ *vs.* $\mathbf{N}_j$ is given by (9.6). As an illustration, for $(i, j) = (4, 3)$, the LR statistic λ_{43} is given by

$$\lambda_{43}^{2/n} = \frac{|\hat{\Sigma}_{\{1\}}||\hat{\Sigma}_{\{2,3\}\bullet\{1\}}||\hat{\Sigma}_{\{4\}\bullet\{1,2,3\}}|}{|\hat{\Sigma}_{0\{1\}}||\hat{\Sigma}_{0\{2\}}||\hat{\Sigma}_{0\{3\}\bullet\{1,2\}}||\hat{\Sigma}_{0\{4\}\bullet\{2,3\}}|}.$$

Example 10. (4-variate two-way MANOVA layout with no interactions and ADG Markov structure.) As in Example 4, let $I = \{1, 2, 3, 4\}$ and now let $N := R \times C$ ($\equiv$ rows $\times$ columns), $n = |R||C|$, so that we observe

$$y \equiv (x_{irc} \mid i \in I,\ r \in R,\ c \in C) \in \mathbf{R}^{I\times R\times C} \equiv \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ x^{(4)} \end{pmatrix},$$

where $x^{(i)} := (x_{irc} \mid r \in R,\, c \in C) \in \mathbf{R}^{R\times C}$. Assume that $\mathrm{E}(y)$ has the usual additive form given by

$$\mathrm{E}(x_{irc}) = \alpha_{ir} + \beta_{ic}, \quad i \in I,\, r \in R,\, c \in C, \tag{13.35}$$

where $\alpha_{ir}, \beta_{ic} \in \mathbf{R}$. Equivalently, $\mathrm{E}(y) \in L$, where

$$L := \mathbf{R}^I \otimes K \subset \mathbf{M}(I\times R\times C)$$

is the MANOVA subspace with

$$K := \bigl\{(\alpha_r + \beta_c \mid r \in R, c \in C) \bigm| (\alpha_r | r \in R) \in \mathbf{R}^R,\, (\beta_c | c \in C) \in \mathbf{R}^C\bigr\} \subset \mathbf{R}^{R\times C}.$$

As in Example 1, let D be the trivial ADG with $V = \{1234\}$ and associated trivial partitioning $I_{1234} \equiv [1234] := \{1, 2, 3, 4\}$, so that L is a D-subspace and $\mathbf{N}_{I\times R\times C}(L, D)$ is a normal linear ADG model, in fact a normal MANOVA model with unrestricted covariance structure. Since $p_L = |R|+|C|-1$ and $|\preceq 1234 \succeq| = 4$, condition (7.2) for the a.e. existence of the MLE is $|R||C| \geq |R| + |C| + 3$.

Suppose that we wish simultaneously to test the hypotheses that $x^{(1)}$ has no row or column effects, $x^{(2)}$ has no column effects, and $x^{(3)}$ has no row effects, i.e.,

$$\mathrm{E}(x_{1rc}) = \gamma, \quad \mathrm{E}(x_{2rc}) = \alpha_r, \quad \mathrm{E}(x_{3rc}) = \beta_c, \quad \mathrm{E}(x_{4rc}) = \alpha'_r + \beta'_c, \quad r \in R,\, c \in C, \tag{13.36}$$

where $\gamma, \alpha_r, \beta_c, \alpha'_r, \beta'_c \in \mathbf{R}$. This combined hypothesis can be expressed equivalently as $\mathrm{E}(y) \in M$, where

$$M := M_{[1]} \times M_{[2]} \times M_{[3]} \times M_{[4]} \subset \mathbf{M}(I\times R\times C)$$

with $M_{[i]} = \mathbf{R}^{[i]} \otimes K_i$, $i = 1, 2, 3, 4$, and where the subspaces $K_i \subset \mathbf{R}^{R\times C}$ are given by

$$\begin{aligned} K_1 &:= \bigl\{(\gamma \mid r \in R, c \in C) \bigm| \gamma \in \mathbf{R}\bigr\},\\ K_2 &:= \bigl\{(\alpha_r \mid r \in R, c \in C) \bigm| (\alpha_r | r \in R) \in \mathbf{R}^R\bigr\},\\ K_3 &:= \bigl\{(\beta_c \mid r \in R, c \in C) \bigm| (\beta_c | c \in C) \in \mathbf{R}^C\bigr\}, \end{aligned}$$

and $K_4 := K$. Since M is not a D-subspace, we cannot test $H_{M,D}$ *vs.* $H_{L,D}$ (recall (9.5) in Remark 9.1) by exact classical methods.

Instead, consider $E := D(M)$, the maximal ADG E such that M is an E-subspace of $\mathbf{M}(I\times R\times C)$, with the associated partitioning of I as in Example 4. Since $K_1 \subset K_2 \subset K_4$ and $K_1 \subset K_3 \subset K_4$, $E \equiv D(M)$ is the TADG in (13.32) with Markov condition $2 \perp\!\!\!\perp 3 \mid 1$. Then $\psi\colon E \to D$ given by $\psi(i) := 1234$, $i = 1, 2, 3, 4$, is a proper surjective ADG homomorphism satisfying (9.1), so we can test exactly either $H_{M,E}$ *vs.* $H_{L,D}$ or $H_{M,E}$ *vs.* $H_{L,E}$ (recall (9.3) and (9.4)). Let $p_i = \dim(K_i)$, $i = 1, 2, 3, 4$. Since $p_1 = 1$, $p_2 = |R| - 1$, $p_3 = |C| - 1$, $p_4 = |R| + |C| - 1$ and, for E, $|\preceq 1 \succeq| = 1$, $|\preceq 2 \succeq| = |\preceq 3 \succeq| = 2$, $|\preceq 4 \succeq| = 4$, the condition in Proposition 9.1 for a.e. existence of the LR statistic for testing $H_{M,E}$ *vs.* $H_{L,E}$ is also $|R||C| \geq |R| + |C| + 3$, the same as for testing $H_{M,E}$ *vs.* $H_{L,D}$. Thus, when this condition holds, by imposing the parsimonious constraint $2 \perp\!\!\!\perp 3 \mid 1$ on the covariance structure we can test (13.35) *vs.* (13.36) by exact classical methods.

Acknowledgement. We wish to thank P. Svante Eriksen, Steffen Lauritzen, and two anonymous referees for helpful discussions and comments.

References

Almond, R. (1995). *Graphical Belief Modelling.* Springer, New York.

Andersen, H. H., Højbjerre, M., Sørensen, D., Eriksen, P. S. (1995). *Linear and Graphical Models for the Multivariate Complex Normal Distribution.* Springer, New York.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

Andersson, S. A. (1990). The lattice structure of orthogonal linear models and orthogonal variance component models. *Scand. J. Statist.* **17** 287-319.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997a). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scand. J. Statist.* **24** 81-102.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997b). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505-541.

Andersson, S. A., Madigan, D., Perlman, M. D., and Triggs, C. M. (1995). On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *J. Statist. Planning and Inference* **48** 25-46.

Andersson, S. A., Madigan, D., Perlman, M. D., and Triggs, C. M. (1997). A graphical characterization of lattice conditional independence models. *Ann. Math. and Artificial Intelligence* **21** 27-50.

Andersson, S. A., Marden, J. I., and Perlman, M. D. (1993). Totally ordered multivariate linear models. *Sankhya Series A* **55** 370-394.

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

Andersson, S. A. and Perlman, M. D. (1994). Normal linear models with lattice conditional independence restrictions. In *Proc. Internat. Symp. Multivariate Analysis Applic.* (T. W. Anderson, K.-T. Fang, I. Olkin, eds.), IMS Lecture Notes-Monograph Series **24** 97-110.

Andersson, S. A. and Perlman, M. D. (1995a). Testing lattice conditional independence models. *J. Multivariate Analysis* **53** 18-38.

Andersson, S. A. and Perlman, M. D. (1995b). Unbiasedness of the likelihood ratio test for lattice conditional independence models. *J. Multivariate Analysis* **53** 1-17.

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* Wiley, New York.

Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statist. Sci.* **8** 204-218; 247-277.

Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis, and Interpretation.* Chapman and Hall, London.

Davey, B. A. and Priestley, H. A. (1990). *Introduction to Lattices and Order.* Cambridge U. Press, Cambridge.

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8** 1272-1317.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28** 157-175.

Eaton, M. L. and Perlman, M. D. (1973). Non-singularity of generalized sample covariance matrices. *Ann. Statist.* **1** 710-717.

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

Eriksen, P. S. (1996). Tests in covariance selection models. *Scand. J. Statist.* **23**

Goldberger, A. S. (1964). *Econometric Theory.* Wiley, New York.

Kiiveri, H., Speed, T. P., and Carlin, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30-52.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford University Press, Oxford.

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

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion.) *J. Royal Statist. Soc. Series B* **50** 157-224.

Lauritzen, S. L. and Wermuth, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31-57.

Ledet Jensen, J. (1991). A large deviation-type approximation for the "Box Class" of likelihood ratio criteria. *J. Amer. Statist. Assoc.* **86** 437-440.

Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems.* Wiley, New York.

Oliver, R. E. and J. Q. Smith, eds. (1990). *Influence Diagrams, Belief Nets, and Decision Analysis.* Wiley, London.

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

Porteous, B. T. (1989). Stochastic inequalities relating a class of log-likelihood ratio statistics to their asymptotic χ^2 distribution. *Ann. Statist.* **17** 1723-1734.

Shachter, R. D. and Kenley, C. R. (1989). Gaussian influence diagrams. *Management Sci.* **35** 527-550.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., and Cowell, R. G. (1993). Bayesian analysis in expert systems (with discussion). *Statist. Sci.* **8** 219-283.

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search.* Springer, New York.

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

Wermuth, N. (1992). On block-recursive regression equations (with discussion). *Brasil. J. Prob. Statist. (Revista Brasileira de Probabilidade e Estatistica)* **6** 1-56.

Wermuth, N. and Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70** 537-552.

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

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57** 348-368.