CONDITIONAL INDEPENDENCE MODELS FOR SEEMINGLY UNRELATED REGRESSIONS WITH INCOMPLETE DATA

MATHIAS DRTON, STEEN A. ANDERSSON, AND MICHAEL D. PERLMAN

ABSTRACT. We consider normal ≡ Gaussian seemingly unrelated regressions (SUR) models with incomplete data (ID). Imposing a natural minimal set of conditional independence constraints, we find restricted SUR/ID models for which the likelihood function and the parameter space factors into the product of the likelihood functions and the parameter spaces of standard complete data multivariate analysis of variance models. Hence, the restricted model has a unimodal likelihood and permits explicit likelihood inference. The restricted model may be used to directly model the data actually observed. Alternatively, the maximum likelihood estimates in the restricted model can yield improved starting values for iterative methods to maximize the likelihood of the unrestricted SUR/ID model. In the development of our methodology, we review and extend existing results for complete data SUR models and the multivariate ID problem. The results are presented in the framework of both lattice conditional independence models and graphical Markov models based on acyclic directed graphs.

Date: October 1, 2003.

Key words and phrases. Acyclic directed graph, graphical model, incomplete data, lattice conditional independence model, MANOVA, maximum likelihood estimator, multivariate analysis, multivariate linear model, missing data, seemingly unrelated regressions.

Contents

1. Introduction

The seemingly unrelated regressions (SUR) model is an extension of the multivariate analysis of variance (MANOVA) model. In MANOVA, each of the observed variables is regressed on the same mean space. In SUR, this is relaxed by allowing different variables to be regressed on different mean spaces. The SUR model was made prominent in the 1960s by Zellner [40, 41] who established the asymptotic efficiency of his two-stage estimator. According to Goldberger [13, p. 323], it "plays a central role in contemporary econometrics". An introduction to SUR can be found, for example, in the econometrics textbooks by Goldberger [13] and by Greene [15] but also in the multivariate statistics monograph by Mardia, Kent, and Bibby [25].

In the SUR literature, several distributional assumptions have been considered. Recent examples include Kowalski et al. [17] who assume t-distributed errors, and Lefkovitch [20] who considers SUR models based on generalized linear models. In this paper, however, we deal exclusively with the classical case of the normal (Gaussian) model.

In general, likelihood inference in a normal SUR model requires iterative methods to maximize the likelihood function (LF). The most common method is the iterated version of Zellner's two-stage estimator; an alternative is Telser's method [36]. Normal SUR models are curved exponential families (van Garderen [37]) and the LF may be multimodal. A bivariate example with multimodal LF is studied in Drton and Richardson [11]. However, in the *monotone* ($\equiv$ *triangular* $\equiv$ *nested*) case in which the regression spaces for the different variables are totally ordered by inclusion the LF is unimodal and explicit likelihood inference is possible by factoring the SUR model into a product of MANOVA models, cf. Andersson and Perlman [8]. Simple special cases can be found earlier, see e.g. Oksanen [30] for a bivariate example.

Andersson and Perlman's methodology [8] also covers *nonmonotone* SUR models. Using *lattice conditional independence* (LCI) theory cf. [6], they show that a nonmonotone SUR model determines a unique minimal set of conditional independence (CI) restrictions s.t. the LCI-restricted nonmonotone SUR model allows for explicit likelihood inference and has a unimodal LF. As in the monotone case, the key idea is the factorization of the SUR model into a product of MANOVA models. In a Monte Carlo study, Wu and Perlman [39] compare the finite sample performance of the estimators obtained in the LCI-restricted SUR model to traditional methods such as ordinary least squares or Zellner's two-stage estimator.

LCI theory also may be applied to incomplete data (ID) problems where, as assumed throughout this article, data is missing at random and the missing data can be ignored in the formulation of the likelihood for the incomplete data (compare Little and Rubin [21, Ch. 6]). For the case of i.i.d. multivariate normal observations with *monotone* incomplete data, explicit likelihood inference is again possible since the LF can be factored s.t. each factor corresponds to a complete data MANOVA model (see Little and Rubin [21, Ch. 7] and Liu [22]). As noted by Murray [29], however, *nonmonotone* incomplete data can lead to a multimodal LF, and iterative methods such as the EM algorithm are needed to find the maximum likelihood estimator (MLE) (cf. Little and Rubin [21], Liu [22], and Schafer [31]).

For the i.i.d. nonmonotone case, Andersson and Perlman [5] applied LCI theory to find a unique minimal set of LCI restrictions s.t. the model again becomes a product of MANOVA models and thus permits explicit likelihood inference. In particular, the LCI-restricted ID model has a unimodal LF. Earlier, Little and Rubin [21, Ch. 7] had introduced the idea of a CI restriction in a simple trivariate example.

These two applications of LCI theory come together in a SUR model with incomplete data—considered for example by Hwang [16], Schmidt [32], Sharma [33], and Swamy and Mehta [35], where the latter work in a Bayesian setting. The LF of the *SUR/ID model* inherits multimodality from both the SUR model and the ID model. Meng and Rubin [26, 27] introduce the ECM algorithm as a generalization of the EM algorithm, in which the M-step is replaced by several conditional or constrained maximization steps. They demonstrate in particular how ECM can be used to fit a SUR/ID model.

In the present paper, we combine the LCI theories developed for SUR and for incomplete data to find a minimal set of LCI restrictions that guarantee a unimodal LF for the SUR/ID model and render explicit likelihood inference possible. We develop the methodology in the LCI framework but also show how, equivalently, the resulting minimal set of CIs can be found using graphical Markov models based on *acyclic digraphs* (ADG $\equiv$ DAG), cf. [2, 3, 9].

The practical value of our results is two-fold. On one hand, the parsimonious LCI model may be reasonable and in good agreement with the data. In this case one avoids the use of iterative methods and the possible difficulty of having to decide which local maximum of the LF yields the most desirable estimate. On the other hand, even if we so not wish to impose CI restrictions, the parsimonious LCI model can be employed to obtain starting values for iterative procedures. These starting values may avoid non-convergence and/or lead to faster convergence than starting values from ordinary least squares, which assumes complete independence of all observed variables. Furthermore, the estimates in the LCI model may help identify the most desirable local maximum if one is confronted with a multimodal LF.

The paper is organized as follows. In Section 2 we introduce LCI theory. Since we aim to find minimal sets of CI restrictions we show in Section 3 how two LCI models can be compared for inclusion. The applications of LCI theory to SUR and to the ID problem are presented in Sections 4 and 5, respectively. The first new result is given in Section 6, where we introduce and study the *linear ID model*. This model includes in particular the MANOVA model with incomplete data. We prove that the parsimonious LCI model for i.i.d. incomplete data (as found in Andersson and Perlman [5]) also allows a factorization of the linear ID model into a product of complete data MANOVA models.

Our main result is presented in Section 7. Here we show how a minimal set of LCI constraints can be found s.t. the SUR/ID model factors into a product of complete data MANOVA models. In Section 8 we show how a transitive ADG (TADG) can be determined by a SUR model, a linear ID model, or a SUR/ID model s.t. the graphical Markov model based on this TADG is equivalent to the parsimonious LCI model developed in Sections 4, 6, or 7, respectively. A series of examples in Section 9 illustrates our methodology. We conclude with a summary and some comments in Section 10.

2. Lattice conditional independence theory

2.1. **Lattices and LCI models.** Let $Y \equiv (Y_i \mid i \in I) \sim \mathcal{N}(\mu, \Sigma)$ be a normal $\equiv$ Gaussian random vector in $\mathbb{R}^I$, where I is a finite index set, $\mu \in \mathbb{R}^I$, and $\Sigma \in \mathbf{P}(I)$ (the cone of all real positive definite $I \times I$ matrices). Let $\mathcal{K}$ be a *ring* of subsets of I, that is, a subset of the power set 2^I closed under intersection and union, hence a *finite distributive lattice*. We always assume that $\emptyset \in \mathcal{K}$ and $I \in \mathcal{K}$. When referring to a *lattice* we will always mean a ring of subsets of I.

The *LCI model* determined by $\mathcal{K}$ places the following CI constraints on the distribution of Y:

$$Y_K \perp\!\!\!\perp Y_L \mid Y_{K\cap L} \quad \forall K, L \in \mathcal{K}, \tag{2.1}$$

or less redundantly,

$$Y_{K\setminus L} \perp\!\!\!\perp Y_{L\setminus K} \mid Y_{K\cap L} \quad \forall K, L \in \mathcal{K}. \tag{2.2}$$

Here, Y_K denotes the subvector $(Y_i \mid i \in K)$ and $\perp\!\!\!\perp$ denotes (conditional) independence. The set of all covariance matrices Σ s.t. Y satisfies the specified CIs is denoted by $\mathbf{P}(\mathcal{K})$.

Likelihood inference for the normal LCI model

$$\mathbf{N}(\mathcal{K}) := (\mathcal{N}(\mu, \Sigma) \mid \mu \in \mathbb{R}^I,\ \Sigma \in \mathbf{P}(\mathcal{K})) \tag{2.3}$$

on $\mathbb{R}^I$ and its extension to the normal linear LCI model (cf. (2.15), Proposition 2.3, and Theorem 2.4) is based on the partially ordered set (poset) $\mathcal{J}(\mathcal{K})$ of *join-irreducible elements* of the lattice $\mathcal{K}$, ordered by inclusion. For $K \in \mathcal{K}$, $K \neq \emptyset$, define

$$\begin{aligned} \langle K \rangle &:= \cup\,(L \in \mathcal{K} \mid L \subsetneq K), \\ [K] &:= K \setminus \langle K \rangle; \end{aligned}$$

thus $K = \langle K \rangle \,\dot\cup\, [K]$. Now define

$$\begin{aligned} \mathcal{J}(\mathcal{K}) :=& \{K \in \mathcal{K} \mid K \neq \emptyset, \langle K \rangle \subsetneq K\} \\ =& \{K \in \mathcal{K} \mid K \neq \emptyset, [K] \neq \emptyset\}. \end{aligned} \tag{2.4}$$

Equivalently, $K \in \mathcal{J}(\mathcal{K})$ iff $K \neq \emptyset$ and $K = L \cup M \Rightarrow K = L$ or $K = M$. By Proposition 2.1 of [6], every set $L \in \mathcal{K}$ can be partitioned as

$$L = \dot\cup([K] \mid K \in \mathcal{J}(\mathcal{K}),\ K \subseteq L). \tag{2.5}$$

In particular, the index set I can be partitioned as

$$I = \dot{\cup}([K] \mid K \in \mathcal{J}(\mathcal{K})), \tag{2.6}$$

and, for $L \in \mathcal{J}(\mathcal{K})$,

$$\langle L \rangle = \dot{\cup}([K] \mid K \in \mathcal{J}(\mathcal{K}),\ K \subsetneq L). \tag{2.7}$$

See Andersson [1], Davey and Priestley [10, Ch. 2 and 5], or Graetzer [14, Ch. II] for further properties of the poset $\mathcal{J}(\mathcal{K})$. In particular, by Birkhoff's Representation Theorem the join-irreducible elements determine the lattice $\mathcal{K}$ uniquely.

2.2. **The algebra of generalized block-triangular matrices with lattice structure.** For two index sets I and J we denote the vector space of $I \times J$ matrices by $\mathbb{R}^{I\times J}$. However, if the set $\mathbb{R}^{I\times J}$ acts on another set of matrices by left multiplication then we denote it by $\mathbf{M}(I \times J)$, or by $\mathbf{M}(I)$ when $I = J$. Further, $A_{I'\times J'}$ denotes the $I' \times J'$ submatrix of a matrix $A \in \mathbf{M}(I \times J)$.

In accordance with the partition (2.6) of the index set I, we can partition a matrix $A \in \mathbf{M}(I)$ as

$$A = (A_{[L]\times[M]} \mid L, M \in \mathcal{J}(\mathcal{K})), \tag{2.8}$$

For each lattice $\mathcal{K}$ define

$$\mathbf{M}(\mathcal{K}) := \{A \in \mathbf{M}(I) \mid A_{[L]\times[M]} = 0 \quad \forall\, M \not\subseteq L \in \mathcal{J}(\mathcal{K})\}. \tag{2.9}$$

It is shown in [6, Sect. 2.4] that $\mathbf{M}(\mathcal{K})$ is an algebra of *generalized block-triangular matrices*. In particular, for $K \in \mathcal{J}(\mathcal{K})$ it follows from (2.7) and (2.9) that the $K \times K$ submatrix A_K of $A \in \mathbf{M}(\mathcal{K})$ has the form

$$A_K = \begin{pmatrix} A_{\langle K\rangle} & 0 \\ A_{[K\rangle} & A_{[K]} \end{pmatrix}, \tag{2.10}$$

where $A_{\langle K\rangle}$, $A_{[K\rangle}$, and $A_{[K]}$ are the $\langle K\rangle \times \langle K\rangle$, $[K] \times \langle K\rangle$, and $[K] \times [K]$ submatrices of A (compare Remark 2.2 in [6]).

The matrices in $\mathbf{M}(\mathcal{K})$ can be characterized alternatively as follows.

Proposition 2.1 (Proposition 2.2 in [6])**.** *A matrix $A \in \mathbf{M}(\mathcal{K})$ iff one of the following two equivalent conditions is fulfilled:*

(i) $\forall y \in \mathbb{R}^I, \forall L \in \mathcal{K} : y_L = 0 \Rightarrow (Ay)_L = 0$;
(ii) $\forall y \in \mathbb{R}^I, \forall L \in \mathcal{K} : (Ay)_L = A_L y_L$.

Condition (i) implies that $\mathbf{M}(\mathcal{K})$ is closed under matrix multiplication and contains the identity, and thus indeed is an algebra. For further details on $\mathbf{M}(\mathcal{K})$, see Andersson and Perlman [6, Sect. 2.4], [7].

2.3. **The MANOVA model.** Let N be a finite index set and assume that we observe the variables indexed by I on subjects indexed by N. Arranged in matrix form, we observe the random array

$$X \sim \mathcal{N}\left(\mu, \Sigma \otimes 1_N\right) \in \mathbb{R}^{I \times N}, \tag{2.11}$$

where $\mathcal{N}$ denotes the normal distribution, 1_N is the $N \times N$ identity matrix, the columns $X_{\cdot j}$, $j \in N$, of X are independent with common covariance matrix $\Sigma \in \mathbf{P}(I)$, and where $\mathrm{E}X \equiv \mu \in \mathbb{R}^{I \times N}$ is the array of means. The classical *MANOVA*[1] *model on* $\mathbb{R}^{I \times N}$ is defined as

$$\mathbf{N}(\mathcal{U}) := (\mathcal{N}(\mu, \Sigma \otimes 1_N) \mid \mu \in \mathcal{U},\ \Sigma \in \mathbf{P}(I)), \tag{2.12}$$

where $\mathcal{U}$ is a *MANOVA subspace of* $\mathbb{R}^{I \times N}$, defined as a linear subspace $\mathcal{U} \subseteq \mathbb{R}^{I \times N}$ s.t.

$$\mathbf{M}(I)\mathcal{U} \subseteq \mathcal{U}. \tag{2.13}$$

Proposition 2.2 (Characterization of MANOVA subspaces [4])**.** *For a subspace* $\mathcal{U} \subseteq \mathbb{R}^{I \times N}$, *the following statements are equivalent:*

(i) $\mathcal{U}$ *is a MANOVA subspace;*
(ii) $\mathcal{U} = U^I \equiv \times(U \mid i \in I)$ *where* U *is a subspace of* $\mathbb{R}^N$*;*
(iii) $\mathcal{U} = \{\gamma Z \mid \gamma \in \mathbf{M}(I \times J)\}$ *for some design matrix* $Z \in \mathbb{R}^{J \times N}$ *and some finite index set* J.

Proof. (iii)⇒(i): Obvious, since $\mathbf{M}(I)\mathbb{R}^{I \times J} = \mathbb{R}^{I \times J}$.

(i)⇒(ii): Let $\mathcal{U}_i$ be the projection of $\mathcal{U}$ onto $\mathbb{R}^{\{i\} \times N}$. By (2.13), $\mathcal{U}$ is invariant under left multiplication by permutation matrices, so $\mathcal{U}_i = U \subseteq \mathbb{R}^N$ for all $i \in I$. Hence, $\mathcal{U} \subseteq \times(\mathcal{U}_i \mid i \in I) \equiv U^I$. Next, let $\tau_j \in \mathbb{R}^N$, $j \in J$, be a basis of U. Then, by definition of U as the image of a projection, there exists $\zeta^{(i,j)} \in \mathcal{U}$ s.t. the ith row of $\zeta^{(i,j)}$ equals τ_j. Multiply $\zeta^{(i,j)}$ on the left by the $I \times I$ matrix having a one as ith diagonal entry and zeroes elsewhere and apply (2.13) to see that the $I \times N$ matrix $\mu^{(i,j)}$ with τ_j as ith row and zero entries elsewhere is an element of $\mathcal{U}$. Since the matrices $\mu^{(i,j)}$, $i \in I$, $j \in J$, span U^I, we obtain that $\mathcal{U} = U^I$.

(ii)⇒(iii): Choose a basis $\tau_j \in \mathbb{R}^N$, $j \in J$, of U. Define Z to be the $J \times N$ matrix with rows $(\tau_j \mid j \in J)$. Then $U = \{\delta Z \mid \delta \in \mathbb{R}^J\}$, so $\mathcal{U} = \{\gamma Z \mid \gamma \in \mathbb{R}^{I \times J}\}$. □

The MLEs in a MANOVA model are available explicitly (compare e.g. Mardia, Kent, and Bibby [25, Ch. 6]):

$$\begin{aligned} \hat{\mu} &= XZ'(ZZ')^{-1}Z, \\ \hat{\Sigma} &= \frac{1}{n}(X - \hat{\mu})(X - \hat{\mu})'. \end{aligned} \tag{2.14}$$

The MLE $(\hat{\mu}, \hat{\Sigma})$ exists a.s. if $|N| \geq |I| + \dim(U)$ where U is the row space of the design matrix Z. Otherwise, the MLE never exists. If it exists the MLE is the unique solution to the likelihood equations.

[1]We do not make a distinction between a multivariate analysis of variance and a multivariate linear regression model.

2.4. **The linear LCI model and its likelihood factorization.** Again assume (2.11). Andersson and Perlman [8] introduced and studied the *linear LCI model on* $\mathbb{R}^{I\times N}$:

$$\mathbf{N}(\mathcal{U},\mathcal{K}) := (\mathcal{N}(\mu,\Sigma\otimes 1_N) \mid \mu\in\mathcal{U},\ \Sigma\in\mathbf{P}(\mathcal{K})) \tag{2.15}$$

where $\mathcal{U}$ is a $\mathcal{K}$-*linear subspace* (or simply $\mathcal{K}$-*subspace*) *of* $\mathbb{R}^{I\times N}$, defined as a linear subspace of $\mathbb{R}^{I\times N}$ that satisfies

$$\mathbf{M}(\mathcal{K})\mathcal{U}\subseteq\mathcal{U}. \tag{2.16}$$

In contrast to a MANOVA model, a linear LCI model restricts the covariance matrix Σ because $\mathbf{P}(\mathcal{K})\subseteq\mathbf{P}(I)$ but, since $\mathbf{M}(\mathcal{K})\subseteq\mathbf{M}(I)$, it allows the mean matrix μ to lie in a more general subspace while still permitting explicit likelihood inference (cf. Theorem 2.4).

Proposition 2.3 (Characterization of $\mathcal{K}$-subspaces [8])**.** *Let $\mathcal{U}$ be a linear subspace of $\mathbb{R}^{I\times N}$. For each $K\in\mathcal{J}(\mathcal{K})$ let $\mathcal{U}_{[K]}$ and $\mathcal{U}_{\langle K\rangle}$ denote the projections of $\mathcal{U}$ onto $\mathbb{R}^{[K]\times N}$ and $\mathbb{R}^{\langle K\rangle\times N}$, respectively. Then $\mathcal{U}$ is a $\mathcal{K}$-subspace of $\mathbb{R}^{I\times N}$ iff the following three conditions are satisfied:*

(i) $\mathcal{U}=\times(\mathcal{U}_{[K]}\mid K\in\mathcal{J}(\mathcal{K}))$;
(ii) $\forall K\in\mathcal{J}(\mathcal{K})$, $\mathcal{U}_{[K]}$ *is a MANOVA subspace of* $\mathbb{R}^{[K]\times N}$;
(iii) $\forall K\in\mathcal{J}(\mathcal{K})$, $\mathbf{M}([K]\times\langle K\rangle)\mathcal{U}_{\langle K\rangle}\subseteq\mathcal{U}_{[K]}$.

Proof. Theorem 4.2 in [8]; compare also Proposition 6.1. □

Under the linear LCI model $\mathbf{N}(\mathcal{U},\mathcal{K})$, the LF factors according to the partitioning (2.6) (cf. [8]), as follows. For $K\in\mathcal{K}$, let X_K (resp., μ_K) denote the $K\times N$ sub-matrix of X (resp., μ) and let Σ_K denote the $K\times K$ sub-matrix of Σ (cf. (2.11)). Partition X_K, μ_K, and Σ_K according to the decomposition $K=\langle K\rangle\,\dot\cup\,[K]$:

$$X_K=\begin{pmatrix}X_{\langle K\rangle}\\ X_{[K]}\end{pmatrix},\quad \mu_K=\begin{pmatrix}\mu_{\langle K\rangle}\\ \mu_{[K]}\end{pmatrix},\quad \Sigma_K=\begin{pmatrix}\Sigma_{\langle K\rangle} & \Sigma_{\langle K]}\\ \Sigma_{[K\rangle} & \Sigma_{[K]}\end{pmatrix}. \tag{2.17}$$

For each $K\in\mathcal{J}(\mathcal{K})$, the conditional distribution of $X_{[K]}$ given $X_{\langle K\rangle}$ is

$$(X_{[K]}\mid X_{\langle K\rangle})\sim\mathcal{N}_{[K]\times N}\left(\xi_{[K]}+\beta_{[K\rangle}X_{\langle K\rangle},\Lambda_{[K]}\otimes 1_N\right), \tag{2.18}$$

where

$$\begin{aligned}\xi_{[K]}&:=\mu_{[K]}-\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}\mu_{\langle K\rangle},\\ \beta_{[K\rangle}&:=\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1},\\ \Lambda_{[K]}&:=\Sigma_{[K]\cdot\langle K\rangle}=\Sigma_{[K]}-\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}\Sigma_{\langle K]}.\end{aligned} \tag{2.19}$$

(Thus, $\beta_{[K\rangle}$ is the matrix of regression coefficients for $X_{[K]}$ given $X_{\langle K\rangle}$ and $\Lambda_{[K]}\otimes 1_N$ is the conditional covariance matrix.) The family $(\xi_{[K]},\beta_{[K\rangle},\Lambda_{[K]}\mid K\in\mathcal{J}(\mathcal{K}))$ comprises the $\mathcal{K}$-*parameters* of the model $\mathbf{N}(\mathcal{U},\mathcal{K})$.

Theorem 2.4 (Factorization of a linear LCI model [8])**.** *Let $\mathcal{U}$ be a $\mathcal{K}$-subspace of $\mathbb{R}^{I\times N}$. Then the LF for the model $\mathbf{N}(\mathcal{U},\mathcal{K})$ factors as*

$$f_{\mu,\Sigma}(x) = \prod_{K\in\mathcal{J}(\mathcal{K})} f_{\xi_{[K]},\beta_{[K\rangle},\Lambda_{[K]}}(x_{[K]} \mid x_{\langle K\rangle}), \tag{2.20}$$

where $f_{\xi_{[K]},\beta_{[K\rangle},\Lambda_{[K]}}(x_{[K]} \mid x_{\langle K\rangle})$ is the LF for the MANOVA model (2.18) on $\mathbb{R}^{[K]\times N}$. Furthermore, the parameter space factors according to the bijective mapping

$$\begin{aligned} \mathcal{U}\times\mathbf{P}(\mathcal{K}) &\to \times\Big(\mathcal{U}_{[K]}\times\mathbb{R}^{[K]\times\langle K\rangle}\times\mathbf{P}([K]) \mid K\in\mathcal{J}(\mathcal{K})\Big) \\ (\mu,\Sigma) &\mapsto \quad \big(\xi_{[K]},\beta_{[K\rangle},\Lambda_{[K]} \mid K\in\mathcal{J}(\mathcal{K})\big)\,. \end{aligned} \tag{2.21}$$

Proof. See Theorems 5.1 and 5.2 in [8] (compare also Theorem 3.1 in [6]). □

Theorem 2.4 enables one to find the MLE of (μ,Σ) by first deriving the MLEs of the $\mathcal{K}$-parameters from the usual formulas for the MLEs in a MANOVA model (see (2.14)), then using the reconstruction algorithm ([5, 6] or, in a slightly different appearance, [9]) to reconstruct the MLE of (μ,Σ) from its estimated $\mathcal{K}$-parameters. In particular, the MLE of (μ,Σ) exists and is the unique solution to the likelihood equations for a.e. $x\in\mathbb{R}^{I\times N}$ iff

$$\begin{aligned} |N| &\geq \max\big\{|[K]| + d_K + |\langle K\rangle| \,\big|\, K\in\mathcal{J}(\mathcal{K})\big\} \\ &= \max\big\{|K| + d_K \,\big|\, K\in\mathcal{J}(\mathcal{K})\big\}\,, \end{aligned} \tag{2.22}$$

where d_K is the dimension of $\mathcal{U}_{[K]}$ divided by $|[K]|$. (Equivalently, if $\mathcal{U}_{[K]} = U^{[K]}$ then d_K is the dimension of $U\subseteq\mathbb{R}^N$.)

3. Lattice inclusion

3.1. **An inclusion criterion based on join-irreducible elements.** As shown in the subsequent sections, for nonmonotone SUR models and/or nonmonotone ID models, LCI theory dictates the construction of minimal sets of CI restrictions that render explicit likelihood inference possible. In order to prove the minimality, we need to compare two LCI models based on different lattices. From the definitions in Section 2.1 it is obvious that

$$\mathcal{K}\subseteq\mathcal{L} \Longrightarrow \mathbf{P}(\mathcal{L})\subseteq\mathbf{P}(\mathcal{K}). \tag{3.1}$$

(The converse is false, as seen by the example $\mathcal{K} = \{\emptyset,\{1\},\{1,2\}\}$ and $\mathcal{L} = \{\emptyset,\{1,2\}\}$ over the index set $I = \{1,2\}$.)

In the following sections we will use (3.1) to establish minimality of the CI restrictions imposed by an LCI model. Hence, we need to be able to compare lattices. The lemma presented in this subsection gives a criterion to check whether two lattices are nested by inclusion based on their join-irreducible elements.

For $i\in I$, let

$$K_i := \cap(K\in\mathcal{K} \mid i\in K) \tag{3.2}$$

be the smallest member of $\mathcal{K}$ containing the index i. As shown in the proof of Proposition 2.1 in [6], K_i is join-irreducible and, thus, is the smallest join-irreducible element containing i. In particular,

$$i \in [K_i] \equiv K_i \setminus \langle K_i \rangle. \tag{3.3}$$

Lemma 3.1 (Inclusion of join-irreducible elements)**.** *Let $\mathcal{K}$ and $\mathcal{L}$ be two lattices over the same index set I. Let K_i and L_i be the smallest join-irreducible elements of $\mathcal{K}$ and $\mathcal{L}$, respectively, that contain the index $i \in I$. Then*

$$\mathcal{K} \subseteq \mathcal{L} \iff L_i \subseteq K_i \quad \forall i \in I. \tag{3.4}$$

Moreover, if $L_i \subseteq K_i$ for all $i \in I$ then $[L_i] \subseteq [K_i]$ for all $i \in I$.

Proof. First, apply Lemma 1 of [34] with $\mathcal{S} = \mathcal{K}$, $\mathcal{F}_i = K_i$, $B = K$, where K_i is defined in (3.2), to obtain that

$$K \in \mathcal{K} \iff K = \bigcup_{i \in K} K_i. \tag{3.5}$$

($\Rightarrow$): If $\mathcal{K} \subseteq \mathcal{L}$ then in particular $K_i \in \mathcal{L}$ for all i. But then since $i \in K_i$ it follows from (3.5) with $\mathcal{K}$ replaced by $\mathcal{L}$ that

$$K_i = \bigcup_{j \in K_i} L_j \supseteq L_i.$$

($\Leftarrow$): Since $\mathcal{L}$ is closed under union it suffices by (3.5)to show that all K_i are elements of $\mathcal{L}$. Let

$$K_i(\mathcal{L}) := \cup(L \in \mathcal{J}(\mathcal{L}) \mid L \subseteq K_i) \in \mathcal{L};$$

we shall show that $K_i = K_i(\mathcal{L})$. By its definition, $K_i(\mathcal{L}) \subseteq K_i$. To show $K_i \subseteq K_i(\mathcal{L})$ let $j \in K_i$. Then $K_j \subseteq K_i$, since $K_i \in \mathcal{J}(\mathcal{K})$ and K_j is the smallest element of $\mathcal{J}(\mathcal{K})$ that contains j. Since $L_j \subseteq K_j$ by assumption, $L_j \subseteq K_i$, hence $j \in L_j \subseteq K_i(\mathcal{L})$.

The claim that $[L_i] \subseteq [K_i]$ can be established as follows. Assume that $j \in [L_i]$. Then since also $j \in [L_j]$ it follows that $j \in [L_i] \cap [L_j]$, and we can deduce from (2.6) that $L_j = L_i$. Hence, $j \in L_j = L_i \subseteq K_i$, from which it follows that $K_j \subseteq K_i$. On the other hand, $i \in L_i = L_j \subseteq K_j$ so $K_i \subseteq K_j$, which implies $K_j = K_i$. In particular, $j \in [K_j] = [K_i]$ which establishes that $[L_i] \subseteq [K_i]$. ∎

3.2. **An inclusion criterion based on the algebra of generalized block-triangular matrices.** The inclusion of two lattices is also characterized by the inclusion of their associated algebras of generalized block-triangular matrices, which we present as Corollary 3.3 to the prepatory Lemma 3.2.

Lemma 3.2. *Let $\mathcal{K}$ be a lattice and define*

$$\tilde{\mathcal{K}} := \{K \subseteq I \mid \forall A \in \mathbf{M}(\mathcal{K}), \forall y \in \mathbb{R}^I :\ y_K = 0 \Rightarrow (Ay)_K = 0\}. \tag{3.6}$$

Then $\tilde{\mathcal{K}} = \mathcal{K}$.

Proof. ($\supseteq$) : Proposition 2.1(i) implies that $\mathcal{K} \subseteq \tilde{\mathcal{K}}$. ($\subseteq$) : We show that $L \notin \mathcal{K} \Rightarrow L \notin \tilde{\mathcal{K}}$. For $L \notin \mathcal{K}$, set $L^- := \cup(K \in \mathcal{K} \mid K \subseteq L) \in \mathcal{K}$ and define $\underline{L} := L \setminus L^- \neq \emptyset$; set $L^+ := \cap(K \in \mathcal{K} \mid K \supseteq L) \in \mathcal{K}$ and define $\overline{L} := L^+ \setminus L \neq \emptyset$. Then for $A \in \mathbf{M}(\mathcal{K})$ and $y \in \mathbb{R}^I$ s.t. $y_L = 0$, we have

$$(Ay)_{\underline{L}} = A_{\underline{L}\times(I\setminus L)} y_{I\setminus L} = A_{\underline{L}\times\overline{L}} y_{\overline{L}} \neq 0. \tag{3.7}$$

In (3.7), the first equality holds since $y_L = 0$, and the second equality is true because $A \in \mathbf{M}(\mathcal{K})$ implies by (2.9) that $A_{(L^+)\times(I\setminus L^+)} = 0 \Rightarrow A_{\underline{L}\times(I\setminus L^+)} = 0$. Finally, $A_{\underline{L}\times\overline{L}} y_{\overline{L}} \neq 0$ if A and y are chosen appropriately, for example, such that $A_{\underline{L}\times\overline{L}}$ and $y_{\overline{L}}$ both have positive entries only. Note that choosing $A_{\underline{L}\times\overline{L}}$ with positive entries does not contradict $A \in \mathbf{M}(\mathcal{K})$ since there exists $K \in \mathcal{J}(\mathcal{K})$ s.t. $\underline{L} \dot\cup \overline{L} = [K]$ (compare (2.9)). In conclusion, $(Ay)_L \neq 0$ and thus $L \notin \tilde{\mathcal{K}}$. □

Corollary 3.3 (Inclusion of matrix algebras)**.** *Let $\mathcal{K}$ and $\mathcal{L}$ be two lattices over the same index set I with associated algebras of generalized block-triangular matrices $\mathbf{M}(\mathcal{K})$ and $\mathbf{M}(\mathcal{L})$, respectively. Then*

$$\mathcal{K} \subseteq \mathcal{L} \iff \mathbf{M}(\mathcal{L}) \subseteq \mathbf{M}(\mathcal{K}). \tag{3.8}$$

Proof. ($\Rightarrow$): By Proposition 2.1(ii), $A \in \mathbf{M}(\mathcal{L})$ iff

$$(Ay)_L = A_L y_L \quad \forall L \in \mathcal{L}. \tag{3.9}$$

Since $\mathcal{K} \subseteq \mathcal{L}$ by assumption, (3.9) holds for all $L \in \mathcal{K}$, hence $A \in \mathbf{M}(\mathcal{K})$ (compare Andersson and Perlman [7, Equation (2.3)]).

($\Leftarrow$): The inclusion $\mathbf{M}(\mathcal{L}) \subseteq \mathbf{M}(\mathcal{K})$ implies that $\tilde{\mathcal{K}} \subseteq \tilde{\mathcal{L}}$, hence $\mathcal{K} \subseteq \mathcal{L}$ by Lemma 3.2. □

4. Seemingly unrelated regressions

4.1. **The SUR model.** The general normal SUR model on $\mathbb{R}^{I\times N}$ is determined by a *SUR pair* $\mathbb{S} := (\mathfrak{U}, I_{\mathfrak{U}})$ *for* $\mathbb{R}^{I\times N}$. Here, the *SUR pattern* $\mathfrak{U}$ is a collection of distinct subspaces of $\mathbb{R}^N$ with $|\mathfrak{U}| \leq |I|$ and the *SUR partition*

$$I_{\mathfrak{U}} := (I_U \mid U \in \mathfrak{U}) \tag{4.1}$$

is a partition of I indexed by $\mathfrak{U}$, i.e. $I_{\mathfrak{U}}$ is a family of non-empty pairwise disjoint subsets of I s.t.

$$I = \dot\cup(I_U \mid U \in \mathfrak{U}). \tag{4.2}$$

The *SUR linear subspace* (or simply *SUR subspace*) *of* $\mathbb{R}^{I\times N}$ *induced by* $\mathbb{S}$ is defined as

$$\mathcal{U}_{\mathbb{S}} := \times(U^{I_U} \mid U \in \mathfrak{U}) \subseteq \mathbb{R}^{I\times N}. \tag{4.3}$$

The general *normal SUR model on* $\mathbb{R}^{I\times N}$ is defined to be

$$\mathbf{N}(\mathcal{U}_{\mathbb{S}}) := (\mathcal{N}(\mu, \Sigma\otimes 1_N) \mid \mu\in\mathcal{U}_{\mathbb{S}},\ \Sigma\in\mathbf{P}(I)). \tag{4.4}$$

Less formally, in a SUR model each of the variables with index $i \in I_U$ is regressed on the *same* linear regression subspace $U \subseteq \mathbb{R}^N$, so these variables together follow the MANOVA model $\mathbf{N}(U^{I_U})$ on $\mathbb{R}^{I_U\times N}$. These $|\mathfrak{U}|$ MANOVA models are only *seemingly* unrelated because the variables in I_U and $I_{U'}$ may be correlated if Σ is not diagonal. Note also that if $|\mathfrak{U}| = 1$, then $\mathcal{U}_{\mathbb{S}}$ is a MANOVA subspace.

If the SUR pattern $\mathfrak{U}$ is totally ordered with respect to inclusion, the SUR model is called *nested*, or *monotone*. If inclusion yields only a partial ordering of the regression spaces, the SUR model given by (4.4) is called *nonnested*, or *nonmonotone*.

In a nonmonotone normal SUR model, the MLE of (μ, Σ) cannot be found explicitly; instead, iterative methods are required. Furthermore, Drton and Richardson [11] show that the LF may be multimodal and that the standard iterative methods may converge to different local maxima depending upon which starting value is used.

4.2. **LCI restrictions for a SUR model.** For any lattice $\mathcal{K} \subseteq 2^I$, we can impose the associated LCI restrictions on the SUR model $\mathbf{N}(\mathcal{U}_{\mathbb{S}})$ on $\mathbb{R}^{I\times N}$ to obtain the *LCI-restricted SUR model on* $\mathbb{R}^{I\times N}$:

$$\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}) := (\mathcal{N}(\mu, \Sigma\otimes 1_N) \mid \mu\in\mathcal{U}_{\mathbb{S}},\ \Sigma\in\mathbf{P}(\mathcal{K})). \tag{4.5}$$

As shown by Andersson and Perlman [8], $\mathbb{S}$ determines a *unique minimal* lattice $\mathcal{K}_{\mathbb{S}} \subseteq 2^I$ of subsets of I s.t. $\mathcal{U}_{\mathbb{S}}$ becomes a $\mathcal{K}_{\mathbb{S}}$-subspace of $\mathbb{R}^{I\times N}$, hence $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{S}})$ becomes a linear LCI model on $\mathbb{R}^{I\times N}$ amenable to explicit normal-theory likelihood inference, cf. Theorem 2.4. Minimality of $\mathcal{K}_{\mathbb{S}}$ (cf. Theorem 4.3) implies that the corresponding set of LCI constraints is minimal $\equiv$ parsimonious. In this subsection, we will review the construction of $\mathcal{K}_{\mathbb{S}}$ and give an alternate proof of its minimality using the lattice inclusion Lemma 3.1. This alternate proof is more easily adapted to the case of incomplete data considered below.

The set of subspaces $\mathfrak{U}$ is partially ordered by inclusion. For $U \in \mathfrak{U}$, define

$$K_U = \dot\cup(I_{U'} \mid U' \subseteq U,\ U' \in \mathfrak{U}), \tag{4.6}$$

so $K_U \subseteq K_{U'}$ iff $U \subseteq U'$ and $K_U = K_{U'}$ iff $U = U'$. Thus the sets

$$\mathcal{P}_{\mathbb{S}} := \{K_U \mid U \in \mathfrak{U}\} \tag{4.7}$$

and $\mathfrak{U}$ are in 1-1 correspondence and form isomorphic posets under inclusion. Note that $\mathcal{P}_{\mathbb{S}}$ is totally ordered by inclusion iff the SUR pattern is monotone.

The *SUR lattice* $\mathcal{K}_{\mathbb{S}}$ is defined to be the lattice generated by $\mathcal{P}_{\mathbb{S}}$, i.e. the smallest ring containing each K_U, $U \in \mathfrak{U}$. The Birkhoff Representation Theorem (cf. [10, Theorem 5.12] or [1, Theorem

3.2(ii)]) yields that $\mathcal{P}_{\mathbb{S}} = \mathcal{J}(\mathcal{K}_{\mathbb{S}})$ (also compare [8, Sect. 6]). It also follows easily from the definition of K_U that

$$K_U = \cap(K \in \mathcal{K}_{\mathbb{S}} \mid I_U \subseteq K), \tag{4.8}$$

the smallest join-irreducible element in $\mathcal{K}_{\mathbb{S}}$ that contains I_U (or any $i \in I_U$). Furthermore (compare to (3.3))

$$I_U = [K_U], \quad U \in \mathfrak{U}. \tag{4.9}$$

Theorem 4.1 (The parsimonious linear LCI model)**.** *Under the LCI constraints determined by the SUR lattice $\mathcal{K}_{\mathbb{S}}$, the LCI-restricted SUR model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{S}})$ becomes a linear LCI model on $\mathbb{R}^{I\times N}$.*

Proof. By Proposition 2.2(ii), U^{I_U} is a MANOVA subspace of $\mathbb{R}^{I_U\times N}$, hence $\mathcal{U}_{\mathbb{S}} \equiv \times(U^{I_U} \mid U \in \mathfrak{U})$ fulfills conditions (i) and (ii) of Proposition 2.3. Condition (iii) is evident from the isomorphism of the posets $\mathcal{P}_{\mathbb{S}}$ and $\mathfrak{U}$ under the inclusion orderings. Thus $\mathcal{U}_{\mathbb{S}}$ is a $\mathcal{K}_{\mathbb{S}}$-subspace, as required. ☐

By Theorems 2.4 and 4.1, the model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{S}})$ on $\mathbb{R}^{I\times N}$ factors as a product of MANOVA models. By (2.22), the MLE of (μ, Σ) in this model exists and is the unique solution to the likelihood equations for a.e. $x \in \mathbb{R}^{I\times N}$ iff

$$\begin{aligned} |N| &\geq \max\left\{|K_U| + d_U \mid U \in \mathfrak{U}\right\} \\ &= \max\left\{|\dot\cup(I_{U'} \mid U' \subseteq U)| + d_U \mid U \in \mathfrak{U}\right\}, \end{aligned} \tag{4.10}$$

where d_U is the dimension of U.

Remark 4.2. By (2.16) and Corollary 3.3, it follows that if two lattices are nested as $\mathcal{K} \subseteq \mathcal{L}$ then a $\mathcal{K}$-subspace of $\mathbb{R}^{I\times N}$ is also an $\mathcal{L}$-subspace of $\mathbb{R}^{I\times N}$. Hence, for any lattice $\mathcal{L} \supseteq \mathcal{K}_{\mathbb{S}}$, the model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{L})$ is also a linear LCI model on $\mathbb{R}^{I\times N}$. Recall, however, that the larger lattice induces the CIs from the smaller lattice plus possibly further CIs; cf. (3.1). ☐

4.3. **Minimality of the LCI restrictions for a SUR model.** The next theorem states the minimality of the imposed LCI restrictions in Theorem 4.1, first shown by Andersson and Perlman [6]. Here we give a different proof using the lattice inclusion Lemma 3.1.

Theorem 4.3 (Lattice minimality for a SUR model)**.** *The SUR lattice $\mathcal{K}_{\mathbb{S}}$ is uniquely minimal among all lattices $\mathcal{L}$ over I s.t. $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{L})$ is a linear LCI model on $\mathbb{R}^{I\times N}$.*

Proof. Consider a competing lattice $\mathcal{L}$ s.t. $\mathcal{U}_{\mathbb{S}}$ is an $\mathcal{L}$-subspace. For each $i \in I$, let L_i be the smallest join-irreducible element of $\mathcal{L}$ containing i and let $\mathcal{U}_i$ be the projection of $\mathcal{U}_{\mathbb{S}}$ onto $\mathbb{R}^{\{i\}\times N}$. Then by Proposition 2.3(ii) and (iii) it follows that $\mathcal{U}_j \subseteq \mathcal{U}_i$ whenever $j \in L_i$. Let $U(i)$ be the unique member of $\mathfrak{U}$ s.t. $I_{U(i)}$ contains i. By the definition of $K_{U(i)}$, the inclusion $U(j) = \mathcal{U}_j \subseteq \mathcal{U}_i = U(i)$ implies also that $j \in K_{U(i)}$. Thus, all the join-irreducible elements of $\mathcal{L}$

and $\mathcal{K}_{\mathbb{S}}$ are nested as $L_i \subseteq K_{U(i)}$. Because $K_{U(i)} = K_i$ (the smallest join-irreducible element of $\mathcal{K}_{\mathbb{S}}$ that contains i), the lattice inclusion Lemma 3.1 shows that $\mathcal{K}_{\mathbb{S}} \subseteq \mathcal{L}$. □

5. Multivariate incomplete data

5.1. **ID patterns.** Consider a random array $Y \in \mathbb{R}^{I \times N}$ where the variables and subjects are indexed by I and N, respectively. Now, however, some entries of Y may be missing. The *ID pattern* can be described by a subset $\mathbb{I} \subseteq I \times N$ with the interpretation that $(i, n) \in \mathbb{I}$ iff the variable $i \in I$ is observed on subject $n \in N$. To avoid trivialities, assume that no variable and no subject is entirely missing. The set $\mathbb{I}$ can be represented in two canonical ways.

First, for each $n \in N$ let

$$I(n) := \{i \in I \mid (i, n) \in \mathbb{I}\} \in 2^I \setminus \{\emptyset\} \tag{5.1}$$

denote the set of all variables i that are observed on subject n, and define

$$\mathcal{I} := \{I(n) \mid n \in N\} \subseteq 2^I \setminus \{\emptyset\}. \tag{5.2}$$

The *set* $\mathcal{I}$ thus describes the pattern of partially observed column vectors and is called the *column ID pattern*. For each $K \in \mathcal{I}$ define

$$N_K := I^{-1}(K) \equiv \{n \in N \mid I(n) = K\} \neq \emptyset, \tag{5.3}$$

that is, N_K indexes the subjects for which *exactly* the variables in K are observed. Then the family

$$N_{\mathcal{I}} := (N_K \mid K \in \mathcal{I}) \tag{5.4}$$

constitutes a partition of N, called the *column ID partition*, or simply *column partition*, so that

$$N = \dot{\cup}(N_K \mid K \in \mathcal{I}). \tag{5.5}$$

Second, for each $i \in I$ let

$$N(i) := \{n \in N \mid (i, n) \in \mathbb{I}\} \in 2^N \setminus \{\emptyset\} \tag{5.6}$$

denote the set of all subjects n for which variable i is observed, and define

$$\mathcal{N} := \{N(i) \mid i \in I\} \subseteq 2^N \setminus \{\emptyset\}. \tag{5.7}$$

The *set* $\mathcal{N}$ thus describes the pattern of partially observed row vectors and is called the *row ID pattern*. For each $M \in \mathcal{N}$ define

$$I_M := N^{-1}(M) \equiv \{i \in I \mid N(i) = M\} \neq \emptyset, \tag{5.8}$$

that is, I_M indexes the variables that are observed *exactly* on the subjects in M. Then the family

$$I_{\mathcal{N}} := (I_M \mid M \in \mathcal{N}) \tag{5.9}$$

constitutes a partition of I, called the *row ID partition*, or simply *row partition*, so that

$$I = \dot{\cup}(I_M \mid M \in \mathcal{N}). \tag{5.10}$$

In the literature, $\mathcal{I}$ has been referred to as the *incomplete data pattern* (Andersson and Perlman [5]), the *missing data pattern* (Little and Rubin [21, Sect. 1.2]), or the *missingness pattern* (Schafer [31, p. 16]). The column viewpoint leading to the pattern $\mathcal{I}$ is important when specifying a distributional assumption because we usually want subjects to be independent. In explicit likelihood inference, however, we regress a subset of variables in I on another subset of variables (compare Theorem 2.4), thus the row viewpoint leading to the pattern $\mathcal{N}$ is important for statistical analysis.

A column or row ID pattern that is totally ordered by inclusion is called *monotone* or *nested* (compare e.g. Little and Rubin [21]). Proposition 5.4 below shows that the column ID pattern $\mathcal{I}$ is monotone iff the row ID pattern $\mathcal{N}$ is monotone. If either pattern is not totally ordered by inclusion then we refer to it as *nonmonotone* or *nonnested*.

For statistical analysis, for certain $K \subseteq I$ it will be necessary to consider the set of subjects N_K^+ for which all variables in K are observed together. Formally, for any $K \subseteq I$ we define

$$\begin{aligned} N_K^+ &:= \cap\,(N(i) \mid i \in K) \\ &= \dot{\cup}\,(N_{K'} \mid K' \in \mathcal{I},\ K' \supseteq K). \end{aligned} \tag{5.11}$$

Note that for all $M \in \mathcal{N}$, it holds by definition that

$$N_{I_M}^+ = M. \tag{5.12}$$

5.2. **The ID lattice.** Let $\mathcal{K}_\mathcal{I} \subseteq 2^I$ denote the *column ID lattice* (or simply *column lattice*), that is, the lattice generated in 2^I by the column ID pattern $\mathcal{I}$. The row ID pattern $\mathcal{N} \subseteq 2^N \setminus \{\emptyset\}$ does not directly generate a lattice in 2^I. However, $\mathcal{N}$ is partially ordered by inclusion, so we can proceed as follows (compare to §4.2).

For $M \in \mathcal{N}$, define

$$K_M := \dot{\cup}(I_{M'} \mid M' \in \mathcal{N},\ M' \supseteq M), \tag{5.13}$$

so $K_M \subseteq K_{M'}$ iff $M \supseteq M'$ and $K_M = K_{M'}$ iff $M = M'$. It follows that the sets

$$\mathcal{P}_\mathcal{N} := \{K_M \mid M \in \mathcal{N}\} \tag{5.14}$$

and $\mathcal{N}$ are in 1-1 correspondence and form anti-isomorphic posets under inclusion.

The *row ID lattice* (or simply *row lattice*) $\mathcal{K}_\mathcal{N} \subseteq 2^I$ is now defined to be the lattice generated in 2^I by $\mathcal{P}_\mathcal{N}$. Then as in §4.2,

$$\mathcal{P}_\mathcal{N} = \mathcal{J}(\mathcal{K}_\mathcal{N}). \tag{5.15}$$

Moreover, it follows as in (4.8) and (4.9) that K_M is the smallest join irreducible element of $\mathcal{K}_\mathcal{N}$ containing I_M (or any $i \in I_M$) and that

$$I_M = [K_M]. \tag{5.16}$$

Proposition 5.1 (ID lattice)**.** *The column and row lattices coincide, i.e.* $\mathcal{K}_\mathcal{I} = \mathcal{K}_\mathcal{N}$*, and jointly define the* ID lattice $\mathcal{K}_\mathbb{I} := \mathcal{K}_\mathcal{I} = \mathcal{K}_\mathcal{N}$.

Proof. The smallest join-irreducible element of $\mathcal{K}_\mathcal{I}$ containing $i \in I$ is given by (compare (3.2))

$$K'_i = \cap(K \in \mathcal{I} \mid i \in K), \quad i \in I. \tag{5.17}$$

Now, $j \in K'_i$ iff $i \in K$ implies $j \in K$ for all $K \in \mathcal{I}$, that is, iff variable j is observed on a subject n whenever variable i is observed on n. Thus,

$$j \in K'_i \iff N(i) \subseteq N(j) \iff j \in K_{N(i)} \in \mathcal{P}_\mathcal{N}.$$

Hence $K'_i = K_{N(i)}$, which implies that the lattices $\mathcal{K}_\mathcal{I}$ and $\mathcal{K}_\mathcal{N}$ have the same join-irreducible elements, so $\mathcal{K}_\mathcal{I} = \mathcal{K}_\mathcal{N}$. □

Remark 5.2. As in the preceding proof, (3.2) implies that the join-irreducible elements of $\mathcal{K}_\mathbb{I}$ can be described explicitly as follows. The smallest join-irreducible element containing $i \in I$ consists of all $j \in I$ s.t. if variable i is observed on a subject n, then so is variable j. □

Remark 5.3. By (5.15) and (5.16), the row partition $I_\mathcal{N}$ in (5.9) can be expressed equivalently as

$$I_\mathcal{N} = ([K] \mid K \in \mathcal{J}(\mathcal{K}_\mathbb{I})). \quad \square \tag{5.18}$$

Proposition 5.4 (Monotonicity)**.** *The following conditions are equivalent:*

(i) *the column ID pattern* $\mathcal{I}$ *is monotone;*
(ii) *the row ID pattern* $\mathcal{N}$ *is monotone;*
(iii) *the set of join-irreducible elements* $\mathcal{J}(\mathcal{K}_\mathbb{I})$ *is totally ordered by inclusion;*
(iv) *the ID lattice* $\mathcal{K}_\mathbb{I}$ *is totally ordered by inclusion.*

Proof. (iv)⇒(i): Obvious, since $\mathcal{I}$ generates $\mathcal{K}_\mathbb{I}$, and hence $\mathcal{I} \subseteq \mathcal{K}_\mathbb{I}$.

(i)⇒(iii): Equation (5.17) shows that all join-irreducible elements in $\mathcal{J}(\mathcal{K}_\mathbb{I})$ are intersections of elements of $\mathcal{I}$. Since we assume $\mathcal{I}$ to be monotone it follows that $\mathcal{J}(\mathcal{K}_\mathbb{I}) = \mathcal{I}$. Hence, $\mathcal{J}(\mathcal{K}_\mathbb{I})$ is totally ordered by inclusion.

(iii)⇒(iv): By (3.5), every element of a lattice is a union of join-irreducible elements, which implies that $\mathcal{K}_\mathbb{I}$ is totally ordered if $\mathcal{J}(\mathcal{K}_\mathbb{I})$ is.

(ii)⇔(iii): This follows immediately from $\mathcal{P}_\mathcal{N} = \mathcal{J}(\mathcal{K}_\mathcal{N}) = \mathcal{J}(\mathcal{K}_\mathbb{I})$ and the definition of the sets K_M comprised by $\mathcal{P}_\mathcal{N}$ (see (5.13) and (5.14)). □

In the development of LCI theory for ID models we will mainly adopt a column view, but for likelihood inference only the join-irreducible elements $\mathcal{J}(\mathcal{K}_{\mathbb{I}})$ are required. Since we showed that $\mathcal{J}(\mathcal{K}_{\mathbb{I}}) = \mathcal{P}_{\mathcal{N}}$, we need only to determine the row ID pattern $\mathcal{N}$ and the row partition $I_{\mathcal{N}}$ to be able to construct quickly the join-irreducible elements $\mathcal{P}_{\mathcal{N}}$.

6. Linear incomplete data models

6.1. **Linear ID subspaces.** Continuing the discussion from Section 5, suppose now that the complete data array satisfies

$$Y \sim \mathcal{N}(\nu, \Sigma \otimes 1_N), \tag{6.1}$$

where $\nu \in \mathbb{R}^{I\times N}$ and $\Sigma \in \mathbf{P}(I)$. Let X denote the *ID array*, that is, the observed part of Y. By the definition of $\mathbb{I}$, the sample space for X is the vector space $\mathbb{R}^{\mathbb{I}}$, which can be written in the equivalent forms

$$\begin{aligned} \mathbb{R}^{\mathbb{I}} &\equiv \times(\mathbb{R}^{K\times N_K} \mid K \in \mathcal{I}) \\ &= \times(\mathbb{R}^{I_M \times M} \mid M \in \mathcal{N}). \\ &\equiv \times(\mathbb{R}^{[K]\times N_K^+} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})), \end{aligned} \tag{6.2}$$

where the last equivalence follows from (5.12), (5.16), and (5.18). The projection of the complete data array Y onto the ID array X is denoted by

$$\begin{aligned} p_{\mathbb{I}} : \mathbb{R}^{I\times N} &\to \mathbb{R}^{\mathbb{I}}, \\ Y &\mapsto X := (X_K \,|\, K \in \mathcal{I}), \end{aligned} \tag{6.3}$$

where X_K is the $K \times N_K$ submatrix of Y. Then X satisfies

$$X \equiv (X_K \mid K \in \mathcal{I}) \sim \otimes\left(\mathcal{N}\left(\mu_K, \Sigma_K \otimes 1_{N_K}\right) \mid K \in \mathcal{I}\right) \in \mathbb{R}^{\mathbb{I}}, \tag{6.4}$$

where μ_K denotes the $K \times N_K$ submatrix of ν and Σ_K is the $K \times K$ submatrix of Σ. Here

$$E[X] = \mu \equiv (\mu_K \mid K \in \mathcal{I}) := p_{\mathbb{I}}(\nu) \in \mathbb{R}^{\mathbb{I}}. \tag{6.5}$$

In §2.3 and §2.4, linear hypotheses about the mean array μ were given by MANOVA subspaces and $\mathcal{K}$-subspaces defined by invariance under left multiplication by the naturally associated matrix algebras $\mathbf{M}(I)$ and $\mathbf{M}(\mathcal{K})$, respectively. To define an analogous class of subspaces in the ID case, we define the multiplication of an ID array $x \in \mathbb{R}^{\mathbb{I}}$ by a matrix $A \in \mathbf{M}(I)$ as follows (cf. Andersson et al. [4]):

$$Ax := (A_K x_K \mid K \in \mathcal{I}) \in \mathbb{R}^{\mathbb{I}}, \tag{6.6}$$

where A_K is the $K \times K$ submatrix of A. For the linear ID model introduced in §6.2, under appropriate LCI covariance restrictions, explicit likelihood inference is possible for the more general linear hypothesis determined by a *linear ID subspace* (or simply $\mathcal{K}_{\mathbb{I}}$*-subspace*) $\mathcal{U}$ of $\mathbb{R}^{\mathbb{I}}$, that is, a linear subspace of $\mathbb{R}^{\mathbb{I}}$ that fulfills (compare (2.13) and (2.16))

$$\mathbf{M}(\mathcal{K}_{\mathbb{I}})\mathcal{U} \subseteq \mathcal{U}, \tag{6.7}$$

where $\mathcal{K}_{\mathbb{I}}$ is the ID lattice defined in Proposition 5.1.

Proposition 6.1 (Characterization of $\mathcal{K}_{\mathbb{I}}$-subspaces of $\mathbb{R}^{\mathbb{I}}$)**.** *Let $\mathcal{U}$ be a linear subspace of $\mathbb{R}^{\mathbb{I}}$. For each $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$, let $\mathcal{U}^{++}_{[K]}$ and $\mathcal{U}^{++}_{\langle K\rangle}$ denote the projections of $\mathcal{U}$ onto $\mathbb{R}^{[K]\times N^+_K}$ and $\mathbb{R}^{\langle K\rangle\times N^+_K}$, respectively. Then $\mathcal{U}$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$ iff the following three conditions are satisfied:*

(i) $\mathcal{U} = \times(\mathcal{U}^{++}_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}))$;
(ii) $\forall K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$, $\mathcal{U}^{++}_{[K]}$ *is a* MANOVA subspace *of* $\mathbb{R}^{[K]\times N^+_K}$;
(iii) $\forall K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$, $\mathbf{M}([K]\times\langle K\rangle)\mathcal{U}^{++}_{\langle K\rangle} \subseteq \mathcal{U}^{++}_{[K]}$.

Proof. ($\Rightarrow$): Let $A \in \mathbf{M}(\mathcal{K}_{\mathbb{I}})$ and $\mu \in \mathcal{U}$, so by (6.7), $A\mu \in \mathcal{U}$. For $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$, the $K \times N^+_K$ submatrix of $A\mu$ is given by

$$\begin{aligned} (A\mu)^+_K &\overset{(5.11)}{=} \big((A\mu)_{K\times N_{K'}} \mid K' \in \mathcal{I}, K' \supseteq K\big) \\ &\overset{(6.6)}{=} \big(A_K \mu_{K\times N_{K'}} \mid K' \in \mathcal{I}, K' \supseteq K\big) \\ &= A_K\mu^+_K, \end{aligned} \tag{6.8}$$

where the last equality follows from Proposition 2.1(ii). This implies that

$$\begin{aligned} [\mathbf{M}(\mathcal{K}_{\mathbb{I}})\mathcal{U}]^+_K &\overset{(6.8)}{=} \mathbf{M}(\mathcal{K}_{\mathbb{I}})_K\mathcal{U}^+_K \\ &\overset{(2.10)}{=} \begin{pmatrix} \mathbf{M}(\langle K\rangle) & 0 \\ \mathbf{M}([K]\times\langle K\rangle) & \mathbf{M}([K]) \end{pmatrix} \begin{pmatrix} \mathcal{U}^{++}_{\langle K\rangle} \\ \mathcal{U}^{++}_{[K]} \end{pmatrix} \\ &\overset{(6.7)}{\subseteq} \begin{pmatrix} \mathcal{U}^{++}_{\langle K\rangle} \\ \mathcal{U}^{++}_{[K]} \end{pmatrix} \\ &= \mathcal{U}^+_K. \end{aligned}$$

Thus,

$$[\mathbf{M}(\mathcal{K}_{\mathbb{I}})\mathcal{U}]^{++}_{[K]} = \Big[\mathbf{M}([K]\times\langle K\rangle)\mathcal{U}^{++}_{\langle K\rangle}\Big] + \Big[\mathbf{M}([K])\mathcal{U}^{++}_{[K]}\Big] \subseteq \mathcal{U}^{++}_{[K]}, \tag{6.9}$$

which yields (ii) and (iii).

By the definition of $\mathcal{U}^{++}_{[K]}$, $\mathcal{U} \subseteq \times(\mathcal{U}^{++}_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}))$. For any $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$, choose a basis of $\mathcal{U}^{++}_{[K]}$ and let τ be a member of this basis. Since $\mathcal{U}^{++}_{[K]}$ is the projection of $\mathcal{U}$ onto $\mathbb{R}^{[K]\times N^+_K}$, there exists $\zeta \in \mathcal{U}$ s.t. $\zeta^{++}_{[K]} = \tau$. Multiplying ζ by the matrix in $\mathbf{M}(\mathcal{K}_{\mathbb{I}})$ that has the identity matrix $1_{[K]}$ in the $[K]$-th diagonal block and zeroes elsewhere shows that the element $\mu \in \mathbb{R}^{\mathbb{I}}$ with $\mu^{++}_{[K]} = \tau$ and zeroes elsewhere is an element of $\mathcal{U}$. Repeating this for all $K \in \mathcal{J}(\mathcal{K})$ and all basis elements τ shows that $\mathcal{U}$ contains a basis of $\times(\mathcal{U}^{++}_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}))$, so (i) follows.

($\Leftarrow$): If (ii) and (iii) are satisfied then the inclusion in (6.9) holds. This yields that

$$\begin{aligned}\mathbf{M}(\mathcal{K}_{\mathbb{I}})\mathcal{U} &\subseteq \times\left([\mathbf{M}(\mathcal{K}_{\mathbb{I}})\mathcal{U}]^{++}_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})\right)\\ &\overset{(6.9)}{\subseteq} \times\left(\mathcal{U}^{++}_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})\right)\\ &\overset{(i)}{=} \mathcal{U},\end{aligned}$$

hence $\mathcal{U}$ satisfies (6.7). □

Corollary 6.2 (Restriction to a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$)**.** *The space $\mathcal{U}$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$ iff there exists a $\mathcal{K}_{\mathbb{I}}$-subspace $\mathcal{V}$ of $\mathbb{R}^{I\times N}$ s.t. $\mathcal{U} = p_{\mathbb{I}}(\mathcal{V})$. In particular, if $\mathcal{V}$ is a MANOVA subspace of $\mathbb{R}^{I\times N}$ then $p_{\mathbb{I}}(\mathcal{V})$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$.*

Proof. ($\Leftarrow$) : If $\mathcal{V}$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I\times N}$ then Proposition 2.3 implies that $\mathcal{U} := p_{\mathbb{I}}(\mathcal{V})$ fulfills the conditions in Proposition 6.1, thus $\mathcal{U}$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$.

($\Rightarrow$) : Let $\mathcal{U}$ be a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$, which can be written according to Proposition 6.1 as

$$\mathcal{U} = \times(\mathcal{U}^{++}_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})). \tag{6.10}$$

Define

$$\mathcal{V}_{[K]} := \mathcal{U}^{++}_{[K]} \times \mathbb{R}^{[K]\times(N\setminus N^+_K)} \subseteq \mathbb{R}^{[K]\times N}, \tag{6.11}$$

and let

$$\mathcal{V} := \times(\mathcal{V}_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})) \subseteq \mathbb{R}^{I\times N}. \tag{6.12}$$

Then by definition, $p_{\mathbb{I}}(\mathcal{V}) = \mathcal{U}$ and $\mathcal{V}$ satisfies condition (i) of Proposition 2.3. Moreover, $\mathcal{V}$ also fulfills condition (ii) since for all $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$,

$$\begin{aligned}\mathbf{M}([K])\mathcal{V}_{[K]} &= \left(\mathbf{M}([K])\mathcal{U}^{++}_{[K]}\right) \times \left(\mathbf{M}([K])\mathbb{R}^{[K]\times(N\setminus N^+_K)}\right)\\ &\subseteq \left(\mathcal{U}^{++}_{[K]}\right) \times \left(\mathbb{R}^{[K]\times(N\setminus N^+_K)}\right)\\ &= \mathcal{V}_{[K]}.\end{aligned}$$

Here the inclusion is implied by Proposition 6.1(ii). Finally, because $K' \subseteq K \Rightarrow N^+_{K'} \supseteq N^+_K$, it follows that $\mathcal{V}_{\langle K\rangle}$, the projection of $\mathcal{V}$ onto $\mathbb{R}^{\langle K\rangle\times N}$, satisfies

$$\mathcal{V}_{\langle K\rangle} = \times(\mathcal{V}_{[K']} \mid K' \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}),\ K' \subsetneq K) \subseteq \mathcal{U}^{++}_{\langle K\rangle} \times \mathbb{R}^{\langle K\rangle\times(N\setminus N^+_K)}.$$

Thus,

$$\begin{aligned}\mathbf{M}([K]\times\langle K\rangle)\mathcal{V}_{\langle K\rangle} &\subseteq \mathbf{M}([K]\times\langle K\rangle)\left(\mathcal{U}^{++}_{\langle K\rangle} \times \mathbb{R}^{\langle K\rangle\times(N\setminus N^+_K)}\right)\\ &\subseteq \left(\mathbf{M}([K]\times\langle K\rangle)\mathcal{U}^{++}_{\langle K\rangle}\right) \times \left(\mathbf{M}([K]\times\langle K\rangle)\mathbb{R}^{\langle K\rangle\times(N\setminus N^+_K)}\right)\\ &\subseteq \left(\mathcal{U}^{++}_{[K]}\right) \times \left(\mathbb{R}^{[K]\times(N\setminus N^+_K)}\right)\\ &= \mathcal{V}_{[K]},\end{aligned}$$

which shows that condition (iii) of Proposition 2.3 holds and thus that $\mathcal{V}$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I\times N}$. □

Note that this Corollary makes an existence statement only so that if $\mathcal{V} \subseteq \mathbb{R}^{I\times N}$ is not a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I\times N}$ then $p_{\mathbb{I}}(\mathcal{V})$ still might be a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$.

6.2. **LCI restrictions for a linear ID model.** The challenge of maximum likelihood estimation when data is incomplete is somewhat similar to the SUR case. First, if the column or row ID pattern is nonmonotone then the MLE cannot be obtained explicitly. Second, the LF might be multimodal (cf. Murray [29]). But here again, based on the ID pattern/lattice, one can construct a parsimonious LCI model that yields explicit MLEs with guaranteed unimodality of the LF. The original work by Andersson and Perlman [5] treats the case of i.i.d. multivariate normal random vectors. Here, their approach is extended to a linear ID model.

Let $\mathcal{U}$ be a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$. The *linear ID model* $\mathbf{N}(\mathcal{U})$ *on* $\mathbb{R}^{\mathbb{I}}$ is defined as (recall (6.4) and (6.5))

$$\mathbf{N}(\mathcal{U}) := (\otimes (\mathcal{N}(\mu_K, \Sigma_K \otimes 1_{N_K}) \mid K \in \mathcal{I}) \mid \mu \in \mathcal{U},\ \Sigma \in \mathbf{P}(I)). \tag{6.13}$$

For any lattice $\mathcal{K} \subseteq 2^I$, define the *LCI-restricted linear ID model on* $\mathbb{R}^{\mathbb{I}}$:

$$\mathbf{N}(\mathcal{U}, \mathcal{K}) := (\otimes (\mathcal{N}(\mu_K, \Sigma_K \otimes 1_{N_K}) \mid K \in \mathcal{I}) \mid \mu \in \mathcal{U},\ \Sigma \in \mathbf{P}(\mathcal{K})). \tag{6.14}$$

For each $K \in \mathcal{I}$, the $K \times N_K^+$ submatrix X_K^+ of Y is fully observed. Partition X_K^+ as

$$X_K^+ = \begin{pmatrix} X_{\langle K\rangle}^{++} \\ X_{[K]}^{++} \end{pmatrix}, \tag{6.15}$$

so by (6.2)

$$\begin{aligned} X &= (X_K \mid K \in \mathcal{I}) \\ &= (X_{[K]}^{++} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})). \end{aligned} \tag{6.16}$$

Furthermore, let μ_K^+, $\mu_{\langle K\rangle}^{++}$, and $\mu_{[K]}^{++}$ denote the corresponding quantities when $\mu = E[X]$ from (6.5) replaces X in (6.15). Then under the parsimonious LCI-restricted linear ID model $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$, for each $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$ the conditional distribution of $X_{[K]}^{++}$ given $X_{\langle K\rangle}^{++}$ is

$$(X_{[K]}^{++} \mid X_{\langle K\rangle}^{++}) \sim \mathcal{N}_{[K]\times N_K^+}\left(\xi_{[K]}^{++} + \beta_{[K\rangle} X_{\langle K\rangle}^{++}, \Lambda_{[K]} \otimes 1_{N_K^+}\right), \tag{6.17}$$

where

$$\begin{aligned} \xi_{[K]}^{++} &:= \mu_{[K]}^{++} - \Sigma_{[K\rangle} \Sigma_{\langle K\rangle}^{-1} \mu_{\langle K\rangle}^{++}, \\ \beta_{[K\rangle} &:= \Sigma_{[K\rangle} \Sigma_{\langle K\rangle}^{-1}, \\ \Lambda_{[K]} &:= \Sigma_{[K]\cdot\langle K\rangle} = \Sigma_{[K]} - \Sigma_{[K\rangle} \Sigma_{\langle K\rangle}^{-1} \Sigma_{\langle K]}. \end{aligned} \tag{6.18}$$

The family $(\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}))$ comprises the $\mathcal{K}_{\mathbb{I}}$-*parameters* of the model $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$. The following factorization theorem extends the results in Section 3 of [5] from the i.i.d. case to the linear ID case.

Theorem 6.3 (Factorization of the parsimonious LCI-restricted linear ID model)**.** *Let $\mathcal{U}$ be a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$. Then the LF for the model $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$ factors as*

$$f_{\mu,\Sigma}(x) = \prod_{K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})} f_{\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]}} \left(x^{++}_{[K]} \,\middle|\, x^{++}_{\langle K \rangle} \right), \tag{6.19}$$

where $f_{\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]}}\left(x^{++}_{[K]} \mid x^{++}_{\langle K\rangle}\right)$ is the LF of the MANOVA model on $\mathbb{R}^{[K] \times N^+_K}$ given by (6.17). Moreover, the parameter space factors according to the bijective mapping

$$\begin{aligned} \phi_{\mathbb{I}} : \mathcal{U} \times \mathbf{P}(\mathcal{K}_{\mathbb{I}}) &\to \times \left(\mathcal{U}^{++}_{[K]} \times \mathbb{R}^{[K] \times \langle K \rangle} \times \mathbf{P}([K]) \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}) \right) \\ (\mu, \Sigma) &\mapsto \quad \left(\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}) \right). \end{aligned} \tag{6.20}$$

Proof. The factorization (6.19) of the LF follows immediately from the derivation of the fundamental factorization (3.12) in [5]. This derivation does not make use of any structure in the mean matrix μ, in particular, no use is made of the i.i.d. assumption.

The bijectivity of the reparameterization $\phi_{\mathbb{I}}$ can be seen as follows (compare also Proposition 6.1 in [9]). The restricted reparameterization

$$\begin{aligned} \psi_{\mathbb{I}} : \mathbf{P}(\mathcal{K}_{\mathbb{I}}) &\to \left(\beta_{[K\rangle}, \Lambda_{[K]} \,\middle|\, K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}) \right) \\ \Sigma &\mapsto \left(\mathbb{R}^{[K] \times \langle K \rangle} \times \mathbf{P}([K]) \,\middle|\, K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}) \right) \end{aligned} \tag{6.21}$$

is bijective, as proved in Theorem 2.2 in [6]. Thus,

$$\phi_{\mathbb{I}}(\mu, \Sigma) = \phi_{\mathbb{I}}(\mu', \Sigma') \Longrightarrow \psi_{\mathbb{I}}(\Sigma) = \psi_{\mathbb{I}}(\Sigma') \iff \Sigma = \Sigma'.$$

To show that $\phi_{\mathbb{I}}(\mu, \Sigma) = \phi_{\mathbb{I}}(\mu', \Sigma')$ also implies $\mu = \mu'$, choose a never-decreasing listing of the join-irreducible elements $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$, i.e. find $K_1, \ldots, K_q$, $q = |\mathcal{J}(\mathcal{K}_{\mathbb{I}})|$, s.t. $\alpha < \delta \Rightarrow K_\delta \not\subseteq K_\alpha$. Apply the definition (6.18) of $\xi^{++}_{[K_\alpha]}$ successively to find that $\mu^{++}_{[K_\alpha]} = {\mu'}^{++}_{[K_\alpha]}$ for $\alpha = 1, \ldots, q$. This implies $\mu = \mu'$ and hence the injectivity of $\phi_{\mathbb{I}}$.

The surjectivity of $\phi_{\mathbb{I}}$ follows from an augmented version of the reconstruction algorithm in Section 3.3 of [5]. The augmentation consists in replacing the α-th step in which the next part of the mean column vector (denoted $\mu_{[\alpha]}$ in [5]) is obtained by a step in which the next part of the mean matrix is obtained, i.e. by

$$\mu^{++}_{[K_\alpha]} = \xi^{++}_{[K_\alpha]} + \beta_{[K_\alpha\rangle} \mu^{++}_{\langle K_\alpha \rangle}.$$

At the α-th step of the algorithm, $\mu^{++}_{\langle K_\alpha\rangle} \in \mathcal{U}^{++}_{\langle K_\alpha \rangle}$ is constructed. Since $\mathcal{U}$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$ it follows from Proposition 6.1 that

$$\mathbf{M}([K_\alpha] \times \langle K_\alpha \rangle) \mathcal{U}^{++}_{\langle K_\alpha \rangle} \subseteq \mathcal{U}^{++}_{[K_\alpha]}. \tag{6.22}$$

The inclusion (6.22) implies that $\mu^{++}_{[K_\alpha]}$ is constructed to be in $\mathcal{U}^{++}_{[K_\alpha]}$. Since $\mathcal{U} = \times\big(\mathcal{U}^{++}_{[K_\alpha]} \mid \alpha = 1,\dots,q\big)$, the reconstructed μ is an element of $\mathcal{U}$.

For given $\mathcal{K}_{\mathbb{I}}$-parameters $\big(\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]} \,\big|\, K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})\big)$ of $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$ in the claimed image of $\phi_{\mathbb{I}}$, this augmented reconstruction algorithm yields $(\mu, \Sigma) \in \mathcal{U} \times \mathbf{P}(\mathcal{K}_{\mathbb{I}})$ s.t.

$$\phi_{\mathbb{I}}(\mu, \Sigma) = \Big(\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})\Big).$$

Thus, $\phi_{\mathbb{I}}$ is surjective.

Finally, since $\xi^{++}_{[K]}$ ranges through the entire MANOVA subspace $\mathcal{U}^{++}_{[K]}$ of $\mathbb{R}^{[K]\times N^+_K}$, the model (6.17) is a MANOVA model on $\mathbb{R}^{[K]\times N^+_K}$ based on $|N^+_K|$ observations. □

Theorem 6.3 shows that, just as in the SUR case, if the LCI constraints given by $\mathcal{K}_{\mathbb{I}}$ are imposed on the linear ID model $\mathbf{N}(\mathcal{U})$ to produce the LCI-restricted linear ID model $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ then explicit likelihood inference is possible. However, different sample sizes $|N^+_K|$ apply to the different regression factors. Thus, the necessary and sufficient condition for almost sure existence and uniqueness of the MLE in the LCI-restricted linear ID model $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ is that

$$|N^+_K| \ge |K| + d_K \qquad \forall K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}}), \tag{6.23}$$

where d_K is defined in (6.26).

Remark 6.4. Note that a factorization theorem analogous to Theorem 6.3 holds for a LCI-restricted linear ID model $\mathbf{N}(\mathcal{U}, \mathcal{L})$ whenever $\mathcal{L} \supseteq \mathcal{K}_{\mathbb{I}}$. To see this let L_i and K_i denote the smallest join-irreducible elements of $\mathcal{L}$ and $\mathcal{K}_{\mathbb{I}}$, respectively, that contain the index $i \in I$. Then, by the lattice inclusion Lemma 3.1,

$$L_i \subseteq K_i \;\text{ and }\; [L_i] \subseteq [K_i]. \tag{6.24}$$

Hence, the proof of Theorem 6.3 applies since, due to (6.24),

(i) the factorization of the LF of $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ implies the factorization of the LF of $\mathbf{N}(\mathcal{U}, \mathcal{L})$;
(ii) the fact that (6.22) holds for K_α implies that (6.22) remains true if K_α is replaced by L_i, where $i \in I$ is such that $K_i = K_\alpha$;
(iii) for all $i \in I$, $\mathcal{U}^{++}_{[L_i]}$ is a MANOVA subspace because $\mathcal{U}^{++}_{[K_i]}$ is one also. □

6.3. **Minimality of the LCI restrictions for a linear ID model.** The next theorem shows the unique minimality of the lattice $\mathcal{K}_{\mathbb{I}}$, which translates into parsimony of the induced conditional independences.

Theorem 6.5 (Lattice minimality for a linear ID model)**.** *The ID lattice $\mathcal{K}_{\mathbb{I}}$ is uniquely minimal among all lattices $\mathcal{L}$ over I for which the model $\mathbf{N}(\mathcal{U}, \mathcal{L})$ on $\mathbb{R}^{\mathbb{I}}$ admits factorizations of the LF and the parameter space as products of LFs and parameter spaces, respectively, of MANOVA models, as in Theorem 6.3.*

Proof. Let $\mathcal{L}$ be a competing lattice admitting a factorization as in (6.19) and (6.20) in Theorem 6.3 and let L_i be the smallest join-irreducible element of $\mathcal{L}$ containing $i \in I$. Since the L_i-th factor in the factorization must be a MANOVA model, a variable $j \in L_i$ is observed on every subject on which the variable i is observed. However, K_i, the smallest join-irreducible element of $\mathcal{K}_{\mathbb{I}}$ containing the index i, contains all $j \in I$ s.t. variable j is observed whenever variable i is observed (see §5.2). Thus, $L_i \subseteq K_i$ for all $i \in I$. The lattice inclusion Lemma 3.1 then implies that $\mathcal{K}_{\mathbb{I}} \subseteq \mathcal{L}$, hence the minimality and uniqueness of $\mathcal{K}_{\mathbb{I}}$. □

Remark 6.6. The distribution of the random ID array X in (6.4) is uniquely determined by the mean parameter $\mu \in \mathbb{R}^{\mathbb{I}}$ from (6.5) and the covariance matrix Σ. In a linear ID model $\mathbf{N}(\mathcal{U})$ on $\mathbb{R}^{\mathbb{I}}$, the parameter $(\mu, \Sigma) \in \mathcal{U} \times \mathbf{P}(I)$ is identifiable iff Σ is identifiable, which holds iff

(6.25)
$$\cup(K \times K \mid K \in \mathcal{I}) = I \times I;$$

compare [5]. However, if $\mathcal{U} = p_{\mathbb{I}}(\mathcal{V})$ for a $\mathcal{K}_{\mathbb{I}}$-subspace $\mathcal{V}$ of $\mathbb{R}^{I \times N}$ then $\nu \in \mathcal{V}$ need not be uniquely identified by $p_{\mathbb{I}}(\nu)$. For this identifiability to hold, the projection $p_{\mathbb{I}} : \mathcal{V} \to \mathcal{U}$ must be bijective, or equivalently, $\dim(\mathcal{V}) = \dim(\mathcal{U})$. In applications, this condition can be verified as follows.

Since $\mathcal{V}$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I \times N}$, we can write the projection of $\mathcal{V}$ onto $\mathbb{R}^{[K] \times N}$ as $(V_K)^{[K]}$ for a linear subspace $V_K \subseteq \mathbb{R}^N$. Further, let d_K be the dimension of $\mathcal{U}^{++}_{[K]}$ divided by $|[K]|$. Equivalently, if $\mathcal{U}^{++}_{[K]} = (U_K)^{[K]}$ for some univariate linear subspace $U_K \subseteq \mathbb{R}^{N^+_K}$ then

(6.26)
$$d_K = \dim(U_K).$$

Then

(6.27)
$$\begin{aligned} \dim(\mathcal{V}) &= \sum \big(|[K]| \dim(V_K) \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})\big) \\ &\geq \sum \big(|[K]| d_K \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})\big) \\ &= \dim(\mathcal{U}), \end{aligned}$$

so $\dim(\mathcal{V}) = \dim(\mathcal{U})$ iff $d_K = \dim(V_K)$ for all $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$. □

7. Seemingly unrelated regressions with incomplete data

7.1. **The SUR/ID model.** We now combine the ID model considered in Sections 5 and 6 with the SUR model as considered in Section 4. We again observe $X \in \mathbb{R}^{\mathbb{I}}$, normally distributed as in (6.4). Recall from (5.5) and (5.10) that the index sets N and I are partitioned as $(N_K \mid K \in \mathcal{I})$ and $(I_M \mid M \in \mathcal{N})$, respectively. Further, recall from (6.2) that the sample space for X factors as $\mathbb{R}^{\mathbb{I}} = \times(\mathbb{R}^{I_M \times M} \mid M \in \mathcal{N})$.

For each $M \in \mathcal{N}$, let $\mathbb{S}_M \equiv (\mathfrak{U}_M, (I_M)_{\mathfrak{U}_M})$ be a SUR pair for $\mathbb{R}^{I_M \times M}$, that is, the SUR pattern $\mathfrak{U}_M$ is a collection of distinct subspaces of $\mathbb{R}^M$ with $|\mathfrak{U}_M| \leq |I_M|$ and the SUR partition $(I_M)_{\mathfrak{U}_M} \equiv (I_{M,U} \mid U \in \mathfrak{U}_M)$ is a partition of I_M indexed by $\mathfrak{U}_M$, so

(7.1)
$$I_M = \dot\cup(I_{M,U} \mid U \in \mathfrak{U}_M).$$

The *SUR/ID* partition

$$(I_{M,U} \mid M \in \mathcal{N},\ U \in \mathfrak{U}_M) \tag{7.2}$$

of I is as least as fine as the row partition $I_{\mathcal{N}}$. Note that a variable $i \in I_{M,U}$ iff it is observed on all the subjects in M and on no other subject and is regressed on the mean space $U \subseteq \mathbb{R}^M$. The SUR subspace of $\mathbb{R}^{I_M \times M}$ induced by $\mathbb{S}_M$ is given by

$$\mathcal{U}_{\mathbb{S}_M} = \times(U^{I_{M,U}} \mid U \in \mathfrak{U}_M) \subseteq \mathbb{R}^{I_M \times M} \tag{7.3}$$

(recall (4.3)).

The collection

$$\mathbb{S} := (\mathbb{S}_M \mid M \in \mathcal{N}) \tag{7.4}$$

of SUR pairs is called a *SUR/ID structure for* $\mathbb{R}^{\mathbb{I}}$. The *SUR/ID subspace* $\mathcal{U}_{\mathbb{S}}$ *of* $\mathbb{R}^{\mathbb{I}}$ *induced by* $\mathbb{S}$ is defined to be the product space

$$\begin{aligned}\mathcal{U}_{\mathbb{S}} &:= \times\,(\mathcal{U}_{\mathbb{S}_M} \mid M \in \mathcal{N}) \\ &= \times\,(U^{I_{M,U}} \mid M \in \mathcal{N},\ U \in \mathfrak{U}_M).\end{aligned} \tag{7.5}$$

Finally, the *SUR/ID model* $\mathbf{N}(\mathcal{U}_{\mathbb{S}})$ *on* $\mathbb{R}^{\mathbb{I}}$ is defined as (recall (6.4) and (6.5))

$$\mathbf{N}(\mathcal{U}_{\mathbb{S}}) := (\otimes\,(\mathcal{N}\,(\mu_K, \Sigma_K \otimes 1_{N_K}) \mid K \in \mathcal{I}) \mid \mu \in \mathcal{U}_{\mathbb{S}},\ \Sigma \in \mathbf{P}(I)) \tag{7.6}$$

(compare to (4.4) and (6.13)).

Proposition 7.1 (Restriction to a SUR/ID subspace)**.** *The space* $\mathcal{U}$ *is a SUR/ID subspace of* $\mathbb{R}^{\mathbb{I}}$ *iff there exists a SUR subspace* $\mathcal{V}$ *of* $\mathbb{R}^{I\times N}$ *s.t.* $\mathcal{U} = p_{\mathbb{I}}(\mathcal{V})$*.*

Proof. $(\Leftarrow)$: Suppose that $\mathcal{V}$ is a SUR subspace of $\mathbb{R}^{I\times N}$, i.e., $\mathcal{V} = \mathcal{V}_{\mathbb{T}} \subseteq \mathbb{R}^{I\times N}$ where $\mathbb{T} \equiv (\mathfrak{U}, I_{\mathfrak{U}})$ is a SUR pair for $\mathbb{R}^{I\times N}$. Then as in (4.2) and (4.3),

$$\begin{aligned}I &= \dot{\cup}(I_U \mid U \in \mathfrak{U}), \\ \mathcal{V} &= \times(U^{I_U} \mid U \in \mathfrak{U}).\end{aligned}$$

Furthermore, if $\mathcal{V}_M$ denotes the projection of $\mathcal{V}$ onto $\mathbb{R}^{I_M \times M}$ then

$$p_{\mathbb{I}}(\mathcal{V}) = \times(\mathcal{V}_M \mid M \in \mathcal{N}).$$

For all $M \in \mathcal{N}$ and $U \in \mathfrak{U}$, let U_M be the projection of U onto $\mathbb{R}^M$. (Note that it is possible that $U_M = U'_M$ for $U \neq U' \in \mathfrak{U}$.) Each $\mathcal{V}_M$ is a SUR subspace of $\mathbb{R}^{I_M \times M}$ induced by the SUR pair $\mathbb{S}_M \equiv (\mathfrak{U}_m, (I_M)_{\mathfrak{U}_M})$, where $\mathfrak{U}_M = \{U_M \mid U \in \mathfrak{U},\ M \in \mathcal{N},\ I_U \cap I_M \neq \emptyset\}$ with $|\mathfrak{U}_M| \leq |\{(M,U) \mid (M,U) \in \mathcal{N} \times \mathfrak{U},\ I_U \cap I_M \neq \emptyset\}|$. Thus $p_{\mathbb{I}}(\mathcal{V})$ is a SUR/ID subspace of $\mathbb{R}^{\mathbb{I}}$ induced by the SUR/ID structure $\mathbb{S} = (\mathbb{S}_M \mid M \in \mathcal{N})$.

$(\Rightarrow)$: Suppose that $\mathcal{U} = \mathcal{U}_{\mathbb{S}}$ for some SUR/ID structure $\mathbb{S} \equiv ((\mathfrak{U}_M, (I_M)_{\mathfrak{U}_M}) \mid M \in \mathcal{N})$ for $\mathbb{R}^{\mathbb{I}}$. For all $M \in \mathcal{N}$, $U \in \mathfrak{U}_M$, define

$$\begin{aligned} V_{M,U} &:= U \times \mathbb{R}^{N \setminus M} \subseteq \mathbb{R}^N, \\ \mathcal{V} &:= \times(V_{M,U}^{I_{M,U}} \mid M \in \mathcal{N},\ U \in \mathfrak{U}_M) \subseteq \mathbb{R}^{I \times N}. \end{aligned}$$

Let $\mathbb{T}$ be the SUR pair $(\mathfrak{U}, I_{\mathfrak{U}})$ with $\mathfrak{U} = \cup(V_{M,U} \mid M \in \mathcal{N},\ U \in \mathfrak{U}_M)$ and $I_{\mathfrak{U}} = (I_V \mid V \in \mathfrak{U})$, where $I_V := \dot\cup(I_{M,U} \mid V_{M,U} = V)$. Then $p_{\mathbb{I}}(\mathcal{V}) = \mathcal{U}_{\mathbb{S}}$, and $\mathcal{V}$ is a SUR subspace of $\mathbb{R}^{I \times N}$ induced by the SUR pair $\mathbb{T}$. □

7.2. **LCI restrictions for a SUR/ID model.** Let $\mathbb{S} \equiv (\mathbb{S}_M \mid M \in \mathcal{N})$ be a SUR/ID structure for $\mathbb{R}^{\mathbb{I}}$. The SUR/ID model $\mathbf{N}(\mathcal{U}_{\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$ inherits (possible) multimodality of the LF from both the SUR models and the ID problem, as well as the need for iterative methods to find MLEs. Here we combine the LCI theories developed for the SUR case and the ID case to find minimally restrictive LCI constraints that render the LF unimodal and allow explicit determination of the MLE.

Note that for given μ and Σ, the factorization (6.19) of the LF still holds if we impose the LCI restriction $\Sigma \in \mathbf{P}(\mathcal{K}_{\mathbb{I}})$ on $\mathbf{N}(\mathcal{U}_{\mathbb{S}})$. However, we must impose CI restrictions beyond those entailed by $\mathcal{K}_{\mathbb{I}}$ in order to obtain a factorization of the parameter space and to insure that every factor corresponds to a MANOVA model.

For a (complete data) SUR model, the variables in I_U with the common regression space $U \in \mathfrak{U}$ are regressed on the variables in $I_{U'}$ only if the regression space $U' \in \mathfrak{U}$ is a subspace of U, i.e. $U' \subseteq U$. For a linear ID model, the variables in I_M, which are observed on exactly the subjects in $M \in \mathcal{N}$, are regressed on the variables in $I_{M'}$ only if the variables in $I_{M'}$ are always observed with the variables in I_M, i.e. $M' \supseteq M$.

Combining these two ideas motivates a partial ordering on

$$\mathcal{F} := \{(M, U) \mid M \in \mathcal{N},\ U \in \mathfrak{U}_M\} \tag{7.7}$$

as follows. If $M, M' \in \mathcal{N}$ are nested as $M \subseteq M'$ then we can project the space $U \in \mathfrak{U}_{M'}$ onto $\mathbb{R}^M$; denote the image of the projection by U_M^+. Now define

$$(M', U') \leq_{\mathcal{F}} (M, U) \iff \left(M' \supseteq M \text{ and } (U')_M^+ \subseteq U\right). \tag{7.8}$$

The relation $\leq_{\mathcal{F}}$ is a partial ordering:

(i) It is reflexive by definition.

(ii) Moreover, if $(M'', U'') \leq_{\mathcal{F}} (M', U')$ and $(M', U') \leq_{\mathcal{F}} (M, U)$ then $M'' \supseteq M' \supseteq M$. Further, $(U'')_{M'}^+ \subseteq U'$ and $(U')_M^+ \subseteq U$. Since $M' \supseteq M$ it follows that $(U'')_M^+ \subseteq (U')_M^+ \subseteq U$, hence $\leq_{\mathcal{F}}$ is transitive.

(iii) If $(M', U') \leq_{\mathcal{F}} (M, U)$ and $(M, U) \leq_{\mathcal{F}} (M', U')$ then $M = M'$ and $U = U'$. Thus $\leq_{\mathcal{F}}$ is anti-symmetric.

By (7.7), the SUR/ID partition (7.2) can be rewritten as

$$I_{\mathcal{F}} := (I_F \mid F \in \mathcal{F}). \tag{7.9}$$

Now define (compare (4.6) and (5.13))

$$K_F = \dot{\cup}(I_{F'} \mid F' \leq_{\mathcal{F}} F), \tag{7.10}$$

so $K_F \subseteq K_{F'}$ iff $F \leq_{\mathcal{F}} F'$ and $K_F = K_{F'}$ iff $F = F'$. The posets

$$\mathcal{P}_{\mathcal{F}} := \{K_F \mid F \in \mathcal{F}\} \tag{7.11}$$

and $\mathcal{F}$, under inclusion and the partial ordering $\leq_{\mathcal{F}}$ respectively, are isomorphic posets.

We define the *SUR/ID lattice* $\mathcal{K}_{\mathbb{I},\mathbb{S}} \subseteq 2^I$ to be the lattice generated by $\mathcal{P}_{\mathcal{F}}$. In the construction of this lattice, each SUR pair $\mathbb{S}_M \equiv (\mathfrak{U}_M, (I_M)_{\mathfrak{U}_M})$, $M \in \mathcal{N}$, in $\mathbb{S}$ occurs within the corresponding layer $\mathbb{R}^{I_M \times M}$ of the ID sample space $\mathbb{R}^{\mathbb{I}}$. As in §4.2 and §5.2, $\mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}) = \mathcal{P}_{\mathcal{F}}$, K_F is the smallest join irreducible element of $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ containing I_F (or any $i \in I_F$), and $I_F = [K_F]$. (Note that $\mathcal{K}_{\mathbb{I}} \subseteq \mathcal{K}_{\mathbb{I},\mathbb{S}}$.)

For any lattice $\mathcal{K} \subseteq 2^I$, define the *LCI-restricted SUR/ID model on* $\mathbb{R}^{\mathbb{I}}$:

$$\begin{aligned} \mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}) &:= (\otimes (\mathcal{N}(\mu_K, \Sigma_K \otimes 1_{N_K}) \mid K \in \mathcal{I}) \mid \mu \in \mathcal{U}_{\mathbb{S}},\ \Sigma \in \mathbf{P}(\mathcal{K})) \\ &\subseteq \mathbf{N}(\mathcal{U}_{\mathbb{S}}). \end{aligned} \tag{7.12}$$

By Theorem 7.2 below, the parsimonious model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$ factors into a product of MANOVA models. In particular, the SUR/ID subspace $\mathcal{U}_{\mathbb{S}}$ of $\mathbb{R}^{\mathbb{I}}$ is decomposed as a Cartesian product of MANOVA subspaces $[\mathcal{U}_{\mathbb{S}}]^{++}_{[K]}$ that are defined as the projections of $\mathcal{U}_{\mathbb{S}}$ onto $\mathbb{R}^{[K] \times N_K^+}$, $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}})$. Moreover, the K-th MANOVA model, $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}})$, in the factorization in Theorem 7.2 arises from the conditional distribution of $X^{++}_{[K]}$ given $X^{++}_{\langle K \rangle}$, namely

$$(X^{++}_{[K]} \mid X^{++}_{\langle K \rangle}) \sim \mathcal{N}_{[K] \times N_K^+} \left(\xi^{++}_{[K]} + \beta_{[K\rangle} X^{++}_{\langle K \rangle}, \Lambda_{[K]} \otimes I_{N_K^+} \right). \tag{7.13}$$

Here the $\mathcal{K}_{\mathbb{I},\mathbb{S}}$*-parameters* $(\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]} \mid K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}))$ of the model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$ are defined as in (6.18) but with $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ replacing $\mathcal{K}_{\mathbb{I}}$.

Theorem 7.2 (Factorization of the parsimonious LCI-restricted SUR/ID model)**.** *Let* $\mathbb{S} \equiv (\mathbb{S}_M \mid M \in \mathcal{N})$ *be a SUR/ID structure for* $\mathbb{R}^{\mathbb{I}}$*. Then the LF for the model* $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ *on* $\mathbb{R}^{\mathbb{I}}$ *factors as*

$$f_{\mu,\Sigma}(x) = \prod_{K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}})} f_{\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]}} \left(x^{++}_{[K]} \,\middle|\, x^{++}_{\langle K \rangle} \right), \tag{7.14}$$

where $f_{\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]}}\left(x^{++}_{[K]} \,\middle|\, x^{++}_{\langle K \rangle} \right)$ *is the LF of the MANOVA model on* $\mathbb{R}^{[K] \times N_K^+}$ *given by (7.13). The parameter space factors according to the bijective mapping*

$$\begin{aligned} \phi_{\mathbb{I},\mathbb{S}} : \mathcal{U}_{\mathbb{S}} \times \mathbf{P}(\mathcal{K}_{\mathbb{I},\mathbb{S}}) &\to \times \left([\mathcal{U}_{\mathbb{S}}]^{++}_{[K]} \times \mathbb{R}^{[K] \times \langle K \rangle} \times \mathbf{P}([K]) \,\middle|\, K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}) \right) \\ (\mu, \Sigma) &\mapsto \quad \left(\xi^{++}_{[K]}, \beta_{[K\rangle}, \Lambda_{[K]} \,\middle|\, K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}) \right). \end{aligned} \tag{7.15}$$

Proof. The proof of Theorem 6.3 applies with only two steps needing reconsideration. First, we check that for any $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}})$ the subspace $[\mathcal{U}_{\mathbb{S}}]^{++}_{[K]}$ is a MANOVA subspace of $\mathbb{R}^{[K]\times N^+_K}$. By construction of $\mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}})$, the set $[K] = [K_F] = I_F$ for some $F \in \mathcal{F}$. Hence for the unique $M \in \mathcal{N}$ and $U \in \mathfrak{U}_M$ s.t. $(M, U) = F$ it follows from (7.5) that $[\mathcal{U}_{\mathbb{S}}]^{++}_{[K]} = U^{I_F}$, which is a MANOVA subspace $\mathbb{R}^{[K]\times N^+_K}$.

Second, we show that the analogue to inclusion (6.22) holds, i.e. that for $K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}})$,

$$\mathbf{M}([K] \times \langle K\rangle)[\mathcal{U}_{\mathbb{S}}]^{++}_{\langle K\rangle} \subseteq [\mathcal{U}_{\mathbb{S}}]^{++}_{[K]}, \tag{7.16}$$

where $[\mathcal{U}_{\mathbb{S}}]^{++}_{\langle K\rangle}$ is the projection of $\mathcal{U}_{\mathbb{S}}$ onto $\mathbb{R}^{\langle K\rangle\times N^+_K}$. For $i \in K$, let $[\mathcal{U}_{\mathbb{S}}]^{++}_i$ be the projection of $\mathcal{U}_{\mathbb{S}}$ onto $\mathbb{R}^{\{i\}\times N^+_K}$. Then, for $i \in [K]$ and $j \in \langle K\rangle$, (7.10) implies that $[\mathcal{U}_{\mathbb{S}}]^{++}_j \subseteq [\mathcal{U}_{\mathbb{S}}]^{++}_i$, which implies (7.16). □

The necessary and sufficient condition for almost sure existence and uniqueness of the MLE in the model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ is that

$$|N^+_K| \geq |K| + d_K \qquad \forall K \in \mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}), \tag{7.17}$$

where d_K is the dimension of $[\mathcal{U}_{\mathbb{S}}]^{++}_{[K]}$ divided by $|[K]|$. More naturally, since $K = K_F$ for some unique $F = (M, U) \in \mathcal{F}$, $[\mathcal{U}_{\mathbb{S}}]^{++}_{[K]} = U^{I_F}$ and $d_K = \dim(U)$.

Remark 7.3. As in Remark 6.4, an analogue to Theorem 7.2 holds for the model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{L})$ whenever $\mathcal{L} \supseteq \mathcal{K}_{\mathbb{I},\mathbb{S}}$. □

7.3. **Minimality of the LCI restrictions for a SUR/ID model.** The SUR/ID lattice $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ is the unique minimal lattice whose LCI constraints permit factorizations of the forms (7.14) and (7.15) for the LF and parameter space, respectively, in the associated LCI-restricted SUR/ID model.

Theorem 7.4 (Lattice minimality for a SUR/ID model)**.** *The SUR/ID lattice $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ is uniquely minimal among all lattices $\mathcal{L} \subseteq 2^I$ s.t. the LCI-restricted SUR/ID model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{L})$ on $\mathbb{R}^{\mathbb{I}}$ admits a factorization of the LF and the parameter space into a product of LFs and parameter spaces of MANOVA models as in Theorem 7.2.*

Proof. Let $\mathcal{L}$ be a competing lattice permitting factorizations as in (7.14) and (7.15). Let L_i, S_i, and K_i be the smallest join-irreducible elements of the lattices $\mathcal{L}$, $\mathcal{K}_{\mathbb{I}}$, and $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ containing index $i \in I$. The L_i-th factor in the factorization is a MANOVA model on $\mathbb{R}^{[L_i]\times N^+_{L_i}}$. Hence, as in the proof of Theorem 6.5, $L_i \subseteq S_i$ and $N^+_{L_i} = N^+_{S_i}$. Furthermore, $[\mathcal{U}_{\mathbb{S}}]^{++}_{[L_i]}$ is a MANOVA subspace of $\mathbb{R}^{[L_i]\times N^+_{L_i}}$ and, from (7.16), it follows that

$$\mathbf{M}([L_i] \times \langle L_i\rangle)[\mathcal{U}_{\mathbb{S}}]^{++}_{\langle L_i\rangle} \subseteq [\mathcal{U}_{\mathbb{S}}]^{++}_{[L_i]}. \tag{7.18}$$

Now (7.5) yields that $[L_i] \subseteq I_F \subseteq K_F$ for some $F = (M, U) \in \mathcal{F}$ and that $[\mathcal{U}_{\mathbb{S}}]^{++}_{[L_i]} = U^{[L_i]}$. Further, the inclusion (7.18) implies that $\langle L_i\rangle \subseteq \dot\cup(I_{F'} \mid F' \leq_{\mathcal{F}} F) = K_F$. Therefore, $L_i \equiv$

$[L_i]\dot{\cup}\langle L_i\rangle \subseteq K_F$. Since $i \in [L_i] \subseteq I_F$ it follows that $K_F = K_i$ and thus $L_i \subseteq K_i$, for all $i \in I$, which implies $\mathcal{K}_{\mathbb{I},\mathbb{S}} \subseteq \mathcal{L}$ by the lattice inclusion Lemma 3.1. $\square$

Remark 7.5. In the SUR/ID model $\mathbf{N}(\mathcal{U}_{\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$, the parameter $(\mu, \Sigma) \in \mathcal{U}_{\mathbb{S}} \times \mathbf{P}(I)$ is identifiable iff (6.25) holds. Moreover, if $\mathcal{U}_{\mathbb{S}} = p_{\mathbb{I}}(\mathcal{V}_{\mathbb{T}})$ is the restriction of a SUR subspace $\mathcal{V}_{\mathbb{T}}$ of $\mathbb{R}^{I\times N}$ induced by a SUR pair $\mathbb{T} \equiv (\mathfrak{U}, I_{\mathfrak{U}})$ then, as in Remark 6.6, $\nu \in \mathcal{V}_{\mathbb{T}}$ is identified by $\mu = p_{\mathbb{I}}(\nu)$ iff $p_{\mathbb{I}} : \mathcal{V}_{\mathbb{T}} \to \mathcal{U}_{\mathbb{S}}$ is bijective. If we define U_M^+ to be the projection of $U \in \mathfrak{U}$ onto $\mathbb{R}^M$ then $p_{\mathbb{I}}$ is bijective iff $\dim(U_M^+) = \dim(U)$ for all $M \in \mathcal{N}$ and $U \in \mathfrak{U}$ s.t. $I_U \cap I_M \neq \emptyset$.

Remark 7.6. Let $\mathbb{S}$ be a SUR/ID structure for $\mathbb{R}^{\mathbb{I}}$, and let $\mathbb{T} \equiv (\mathfrak{U}, I_{\mathfrak{U}})$ be a SUR pair for $\mathbb{R}^{I\times N}$ s.t. the induced SUR/ID subspace $\mathcal{U}_{\mathbb{S}} \subseteq \mathbb{R}^{\mathbb{I}}$ is the projection onto $\mathbb{R}^{\mathbb{I}}$ of the SUR subspace $\mathcal{V}_{\mathbb{T}} \subseteq \mathbb{R}^{I\times N}$ induced by $\mathbb{T}$ (compare Proposition 7.1). Then one can consider the set

$$\mathcal{F}' := \{(M, U) \mid M \in \mathcal{N},\ U \in \mathfrak{U}\} \tag{7.19}$$

equipped with the partial ordering

$$(M', U') \leq_{\mathcal{F}'} (M, U) \iff (M' \supseteq M \text{ and } U' \subseteq U)\,. \tag{7.20}$$

The lattice induced by this partial ordering equals the lattice $\mathcal{K}(\mathcal{K}_{\mathbb{I}} \cup \mathcal{K}_{\mathbb{T}})$ generated by the union of the ID lattice $\mathcal{K}_{\mathbb{I}}$ and the SUR lattice $\mathcal{K}_{\mathbb{T}}$. However, this lattice is a larger lattice than $\mathcal{K}_{\mathbb{I},\mathbb{S}}$, i.e. $\mathcal{K}(\mathcal{K}_{\mathbb{I}} \cup \mathcal{K}_{\mathbb{T}}) \supseteq \mathcal{K}_{\mathbb{I},\mathbb{S}}$, and in general $\mathcal{K}_{\mathbb{T}} \not\subseteq \mathcal{K}_{\mathbb{I},\mathbb{S}}$ (compare Example 9.6 in Section 9). Recall that smaller lattices lead to more parsimonious models. $\square$

8. Acyclic directed graph theory

8.1. **Directed graphs.** A *graph* is a pair (V, E), where V is a finite set of vertices and $E \subseteq \{(v, w) \in V \times V \mid v \neq w\}$ is a set of edges. An edge $(v, w) \in E$ is undirected if $(w, v) \in E$ and directed if $(w, v) \notin E$. Here we confine ourselves to *directed graphs*, i.e. graphs which contain only directed edges. We denote a (directed) edge in the directed graph $D = (V, E)$ by $v \to w \in D$.

A *path* of length $k \geq 1$ from vertex v to w in D is a sequence of distinct vertices $(v_0, v_1, \dots, v_k)$ s.t. $v_0 = v$, $v_k = w$, and $v_{i-1} \to v_i \in D$ for all $i = 1, \dots, k$. If there exists a path from v to w, we write $v <_D w$ or simply $v < w$. The negation $v \not< w$ denotes that there is no path from v to w. A *cycle* is a path of length $k \geq 3$ from vertex v to itself. A directed graph $D = (V, E)$ is called *acyclic* if it contains no cycles. This is equivalent to $v \not< v$ for all $v \in V$.

In the directed graph $D = (V, E)$, we define the *parents* of a vertex v to be the set $\mathrm{pa}(v) := \{w \in V \mid w \to v \in G\}$. Note that $v \notin \mathrm{pa}(v)$. The *descendants* of v are $\mathrm{de}(v) := \{w \in V \mid v < w\}$. If D is an acyclic directed graph (ADG $\equiv$ DAG) then $v \notin \mathrm{de}(v)$. Finally, the *non-descendants* of v are $\mathrm{nd}(v) := V \setminus (\mathrm{de}(v) \cup \{v\})$.

8.2. **Normal graphical Markov models based on acyclic directed graphs.** Consider a multivariate normal observation $Y = (Y_i \mid i \in I) \sim \mathcal{N}(\mu, \Sigma)$ in $\mathbb{R}^I$ and let $D = (V, E)$ be an ADG, where the vertex set V is itself a partition of I. This means that the vertices in V represent pairwise disjoint sets of variables in I, and that all variables in I are represented. For any subset $A \subseteq V$, Y_A denotes the subvector $(Y_v \mid v \in A)$.

The probability distribution of Y is said to be *D-Markov* if

$$Y_v \perp\!\!\!\perp Y_{\mathrm{nd}(v)\setminus\mathrm{pa}(v)} \mid Y_{\mathrm{pa}(v)} \quad \forall v \in V. \tag{8.1}$$

The set of all $I \times I$ covariance matrices Σ s.t. the CIs specified in (8.1) hold is denoted by $\mathbf{P}(D)$, and the corresponding set of I-variate normal distributions is denoted by

$$\mathbf{N}(D) := (\mathcal{N}(\mu, \Sigma) \mid \mu \in \mathbb{R}^I,\ \Sigma \in \mathbf{P}(D)). \tag{8.2}$$

Note that the smaller the ADG is in the sense of fewer edges, the more CIs it imposes.

Proposition 8.1 (Theorem 1 of [19])**.** *Let f be the density of $Y \sim \mathcal{N}(\mu, \Sigma)$ and let $f(y_v \mid y_{\mathrm{pa}(v)})$ be the conditional density of Y_v given $Y_{\mathrm{pa}(v)}$. Then the distribution of Y is D-Markov iff f factors as*

$$f(y) = \prod_{v \in V} f(y_v \mid y_{\mathrm{pa}(v)}), \quad y \in \mathbb{R}^I. \tag{8.3}$$

More details on the Markov property can be found in Lauritzen et al. [19]. General introductions to graphical Markov models are given by Edwards [12], Lauritzen [18], and Whittaker [38].

8.3. **Equivalence of transitive ADG and LCI models.** The class of LCI models is embedded in the class of ADG models. Andersson et al. [2, 3] show that the class of LCI models is indeed equivalent to the class of *transitive* ADG ($\equiv$ TADG) models. An ADG D is *transitive* if $v \to w \in D$ and $w \to u \in D$ imply $v \to u \in D$. We briefly summarize their equivalence result.

If $\mathcal{K}$ is a lattice then an equivalent TADG $D(\mathcal{K})$ can be constructed as follows. Let $V(\mathcal{K}) := \{[K] \mid K \in \mathcal{J}(\mathcal{K})\}$, which is a decomposition of I (recall (2.6)), and define $E(\mathcal{K})$ by including an edge $[K] \to [K']$ iff $K \subseteq K'$. By the transitivity of set inclusion the constructed graph $D(\mathcal{K}) \equiv (V(\mathcal{K}), E(\mathcal{K}))$ is a TADG, and by [2, Theorem 5.1] and [3, Theorem 4.1], $\mathbf{N}(D(\mathcal{K})) = \mathbf{N}(\mathcal{K})$ (in particular, $\mathbf{P}(D(\mathcal{K})) = \mathbf{P}(\mathcal{K})$).

Conversely, let $D = (V, E)$ be a TADG, where the vertex V is a partition of I. Write $v \leq_D w$ if $v = w$ or $v <_D w$. Since D is transitive it follows that $\leq_D$ is a partial ordering on V. In the spirit of (4.6), (5.13), and (7.10), we set

$$K_v := v \dot{\cup} \{w \mid w \in \mathrm{pa}(v)\} \tag{8.4}$$

for all $v \in V$. Since D is transitive, $K_v \subseteq K_{v'}$ iff $v \leq_D v'$. Hence $\mathcal{P}_V := \{K_v \mid v \in V\}$ and V are isomorphic posets under the inclusion ordering and the partial ordering $\leq_D$, respectively, and the lattice $\mathcal{K}(D)$ generated from $\mathcal{P}_V$ has the join-irreducible elements $\mathcal{J}(\mathcal{K}(D)) = \mathcal{P}_V$.

Then $\mathbf{N}(\mathcal{K}(D)) = \mathbf{N}(D)$, see again [2, Theorem 5.1] and [3, Theorem 4.1], and in particular, $\mathbf{P}(\mathcal{K}(D)) = \mathbf{P}(D)$.

If a lattice $\mathcal{K}$ is monotone, then it does not impose any CIs. This is equivalent to the TADG $D(\mathcal{K})$ being *complete*, i.e. for all vertices v, w either $v \to w \in D(\mathcal{K})$ or $w \to v \in D(\mathcal{K})$.

Andersson and Perlman [9] develop estimation and testing theory in linear regression models when the distribution of Y is normal and D-Markov for an ADG D. In [9, Sect. 10], they show that the larger class of ADG models does not allow a more general linear hypothesis about the mean vector or matrix than the class of TADG models, or equivalently, than the class of LCI models, if the linear ADG model is required to factor into a product of MANOVA models as in our Theorems 2.4, 6.3, and 7.2. Therefore, for our SUR/ID models, consideration of ADG models instead of LCI models will not lead to less restrictive CIs that still allow a MANOVA factorization (compare also the introduction of [3]). However, one can represent the models based on the minimal lattices $\mathcal{K}_{\mathbb{S}}$, $\mathcal{K}_{\mathbb{I}}$, and $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ from Sections 4 through 7 equivalently by TADGs, which sometimes allows an easier graphical representation in applications (see Section 9). This TADG representation is now described.

8.4. **Construction of the TADGs equivalent to the parsimonious LCI models.** For the SUR, incomplete data, and SUR/ID problems, we might first construct the parsimonious lattices $\mathcal{K}_{\mathbb{S}}$, $\mathcal{K}_{\mathbb{I}}$, and $\mathcal{K}_{\mathbb{I},\mathbb{S}}$, respectively, then use the equivalence result in [2, 3] to construct the corresponding TADGs $D(\mathcal{K}_{\mathbb{S}})$, $D(\mathcal{K}_{\mathbb{I}})$, and $D(\mathcal{K}_{\mathbb{I},\mathbb{S}})$. It is simpler, however, to construct the TADGs directly from $\mathbb{S}$, $\mathbb{I}$, and $(\mathbb{I}, \mathbb{S})$, respectively, as follows.

First, as in Section 4, consider a SUR model with SUR pair $\mathbb{S} \equiv (\mathfrak{U}, I_{\mathfrak{U}})$. Take $V_{\mathbb{S}} = I_{\mathfrak{U}}$ (the SUR partition in (4.1)) and define $E_{\mathbb{S}}$ by including the edge $I_U \to I_{U'}$ iff $U \subseteq U'$. Then by (4.6) and (8.4), the resulting TADG $D_{\mathbb{S}} := (V_{\mathbb{S}}, E_{\mathbb{S}})$ satisfies $D_{\mathbb{S}} = D(\mathcal{K}_{\mathbb{S}})$ and $\mathbf{N}(D_{\mathbb{S}}) = \mathbf{N}(\mathcal{K}_{\mathbb{S}})$.

Next, as in Sections 5 and 6, consider an ID problem with ID pattern $\mathbb{I} \subseteq I \times N$. Take $V_{\mathbb{I}} = I_{\mathcal{N}}$ (the row partition in (5.9)) and define $E_{\mathbb{I}}$ by including the edge $I_M \to I_{M'}$ iff $M' \subseteq M$. Then by (5.13) and (8.4), the resulting TADG $D_{\mathbb{I}} := (V_{\mathbb{I}}, E_{\mathbb{I}})$ satisfies $D_{\mathbb{I}} = D(\mathcal{K}_{\mathbb{I}})$ and $\mathbf{N}(D_{\mathbb{I}}) = \mathbf{N}(\mathcal{K}_{\mathbb{I}})$.

Finally, as in Section 7, consider a SUR/ID problem with ID pattern $\mathbb{I} \subseteq I \times N$ and SUR/ID structure $\mathbb{S}$ for $\mathbb{R}^{\mathbb{I}}$. Take $V_{\mathbb{I},\mathbb{S}} = I_{\mathcal{F}}$ (the SUR/ID partition in (7.2) $\equiv$ (7.9)) and define $E_{\mathbb{I},\mathbb{S}}$ by including the edge $I_F \to I_{F'}$ iff $F \leq_{\mathcal{F}} F'$ (recall (7.8)). Then by (7.10) and (8.4), the resulting TADG $D_{\mathbb{I},\mathbb{S}} := (V_{\mathbb{I},\mathbb{S}}, E_{\mathbb{I},\mathbb{S}})$ satisfies $D_{\mathbb{I},\mathbb{S}} = D(\mathcal{K}_{\mathbb{I},\mathbb{S}})$ and $\mathbf{N}(D_{\mathbb{I},\mathbb{S}}) = \mathbf{N}(\mathcal{K}_{\mathbb{I},\mathbb{S}})$.

9. Examples

In this section, we illustrate the LCI and TADG methodology developed herein for SUR models, linear ID models, and SUR/ID models, by a series of examples. In previous work, Andersson and Perlman give an example of LCI theory applied to SUR, see [8, Ex. 6.2], examples of normal linear regression models with ADG covariance structure, see [9, Sect. 13], and examples of the application of LCI theory to incomplete data in the i.i.d. case, see [5, Sect. 4].

In all the examples in this section, we consider four variables indexed by $I := \{1, 2, 3, 4\}$ on the subjects indexed by $N := \{1, \dots, m\}$, with outcomes appearing in the random $I \times N$ matrix Y. Further, assume that the mean matrix $\nu \equiv E[Y]$ is an element of the regression space

$$\Gamma Z \equiv \{\gamma Z \mid \gamma \in \Gamma\} \subseteq \mathbb{R}^{I \times N}. \tag{9.1}$$

Here, Γ is a subspace of $\mathbf{M}(I \times J)$ with $J := \{1, \dots, 5\}$, and Z is the $J \times N$ design matrix

$$Z = \begin{pmatrix} z_{11} & \dots & z_{1m} \\ \vdots & & \vdots \\ z_{51} & \dots & z_{5m} \end{pmatrix} \in \mathbb{R}^{J \times N}, \tag{9.2}$$

which is assumed to be of full rank $|J| = 5 \leq m$. We adopt the notation z_i for the ith row of Z and abbreviate $\{i\}$ by i, $\{i, j\}$ by ij, etc., when no confusion is possible.

The parsimonious lattice and corresponding TADG for the examples are shown in Figures 1 to 6. The lattice is represented by its Hasse diagram (Davey and Priestley [10, Ch. 1], Graetzer [14, Ch. I.2]).

Example 9.1 (*Monotone SUR*)**.** Assume we observe the complete data $X = Y$ with mean matrix $\mu = \nu$ and let $\Gamma \equiv \Gamma_1$ be the subpace of $\mathbf{M}(I \times J)$ given by all matrices of the form

$$\gamma = \begin{pmatrix} \gamma_{11} & 0 & 0 & 0 & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \\ \gamma_{31} & \gamma_{32} & 0 & \gamma_{34} & \gamma_{35} \\ \gamma_{41} & \gamma_{42} & 0 & \gamma_{44} & \gamma_{45} \end{pmatrix}. \tag{9.3}$$

Set

$$U_1 = \operatorname{span}(z_1, z_5), \quad U_2 = \operatorname{span}(z_1, z_2, z_3, z_4, z_5), \quad U_{34} = \operatorname{span}(z_1, z_2, z_4, z_5). \tag{9.4}$$

Define the SUR pattern $\mathfrak{U} := \{U_1, U_2, U_{34}\}$ and let the SUR partition $I_{\mathfrak{U}}$ be given by

$$I_{U_1} = \{1\}, \quad I_{U_2} = \{2\}, \quad I_{U_{34}} = \{3, 4\}, \tag{9.5}$$

Then $\Gamma_1 Z = \mathcal{U}_{\mathbb{S}}$ is the SUR subspace of $\mathbb{R}^{I \times N}$ induced by the SUR pair $\mathbb{S} \equiv (\mathfrak{U}, I_{\mathfrak{U}})$ for $\mathbb{R}^{I \times N}$ (compare Section 4.1).

The regression spaces are ordered as

$$U_1 \subsetneq U_{34} \subsetneq U_2, \tag{9.6}$$

which implies that the join-irreducible elements are $\mathcal{J}(\mathcal{K}_{\mathbb{S}}) = \{1, 134, 1234\}$ and that $\mathcal{K}_{\mathbb{S}} = \{\emptyset, 1, 134, 1234\}$. The graphical representation of $\mathcal{K}_{\mathbb{S}}$ and the equivalent TADG $D_{\mathbb{S}}$ is shown in Figure 1. The SUR lattice $\mathcal{K}_{\mathbb{S}}$ is monotone and the graph $D_{\mathbb{S}}$ is complete, hence no CIs are imposed, i.e. $\mathbf{N}(\mathcal{U}_{\mathbb{S}}) = \mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{S}})$, and the LF for the SUR model $\mathbf{N}(\mathcal{U}_{\mathbb{S}})$ on $\mathbb{R}^{I \times N}$ factors as

$$f(1\ 2\ 3\ 4) = f(2 \mid 1\ 3\ 4)\ f(3\ 4 \mid 1)\ f(1). \tag{9.7}$$

Finally, the condition (4.10) for the almost sure existence and uniqueness of the MLE of (μ, Σ) is fulfilled iff $|N| = m \geq 9 = \max\{1+2, 4+5, 3+4\} = \max\{|K_U| + d_U \mid U \in \mathfrak{U}\}$. $\square$

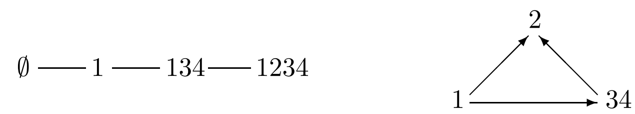

FIGURE 1. The SUR lattice $\mathcal{K}_{\mathbb{S}}$ and corresponding TADG $D_{\mathbb{S}}$.

Example 9.2 (*Nonmonotone SUR*)**.** Change Example 9.1 by setting $\Gamma \equiv \Gamma_2 = \{\gamma \in \Gamma_1 \mid \gamma_{24} = 0\}$, so the regression space of variable 2 changes to $U_2 = \operatorname{span}(z_1, z_2, z_3, z_5)$. If this new U_2 replaces the U_2 in the SUR pattern $\mathfrak{U}$ from Example 9.1 then $\Gamma_2 Z = \mathcal{U}_{\mathbb{S}}$ is the SUR subspace of $\mathbb{R}^{I\times N}$ induced by the SUR pair $\mathbb{S} \equiv (\mathfrak{U}, I_{\mathfrak{U}})$ for $\mathbb{R}^{I\times N}$ where $I_{\mathfrak{U}}$ is the SUR partition from Example 9.1.

The inclusion ordering of the regression spaces changes to

$$U_1 \subsetneq U_2, \quad U_1 \subsetneq U_{34}, \tag{9.8}$$

and the join-irreducible elements are now $\mathcal{J}(\mathcal{K}_{\mathbb{S}}) = \{1, 12, 134\}$. Thus the new lattice $\mathcal{K}_{\mathbb{S}} = \{\emptyset, 1, 12, 134, 1234\}$ is nonmonotone. It imposes the CI $2 \perp\!\!\!\perp 34 \mid 1$, i.e. variable 2 is conditionally independent of variables 3 and 4 given variable 1, and the LF of the LCI-restricted SUR model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{S}})$ on $\mathbb{R}^{I\times N}$ factors as

$$f(1\,2\,3\,4) = f(3\,4 \mid 1)\; f(2 \mid 1)\; f(1). \tag{9.9}$$

The graphical representations of $\mathcal{K}_{\mathbb{S}}$ and $D_{\mathbb{S}}$ are given in Figure 2. Finally, the condition (4.10) for the almost sure existence and uniqueness of the MLE of (μ, Σ) is fulfilled iff $|N| = m \geq 7 = \max\{1+2, 2+4, 3+4\} = \max\{|K_U| + d_U \mid U \in \mathfrak{U}\}$. $\square$

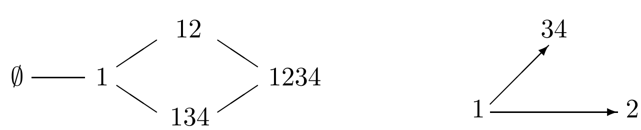

FIGURE 2. The SUR lattice $\mathcal{K}_{\mathbb{S}}$ and corresponding TADG $D_{\mathbb{S}}$.

The remaining Examples 9.3–9.6 have incomplete data with an ID pattern $\mathbb{I}$. With incomplete data our conclusions depend on the exact structure of the subspace $p_{\mathbb{I}}(\Gamma Z)$ of $\mathbb{R}^{\mathbb{I}}$. Hence, we choose $m = 10$ and specify the design matrix

$$
\text{(9.10)} \qquad Z = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array}.
$$

This matrix is of full rank 5 (the rank 5 is checked for a submatrix of Z in Example 9.3a).

Example 9.3a (*MANOVA with monotone incomplete data*)**.** Set $\Gamma \equiv \Gamma_3 = \mathbf{M}(I \times J)$. Then $\Gamma_3 Z = [\text{row}(Z)]^I$ where $\text{row}(Z)$ is the row space of the design matrix Z from (9.10). In particular, $\Gamma_3 Z$ is a MANOVA subspace of $\mathbb{R}^{I \times N}$.

Assume that we observe data as represented by the ID array

$$
\text{(9.11)} \qquad X = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{pmatrix} \blacksquare & & \blacksquare & \blacksquare & \blacksquare & & \blacksquare & & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{pmatrix} \end{array},
$$

where black squares represent the observed values with index in the ID pattern $\mathbb{I} \subseteq I \times N$ and blank spaces represent the missing values. Variable 1 is observed on the subjects in $M_1 := N(1) = \{1, 3, 4, 5, 7, 9, 10\}$, variables 2 and 3 are always observed together and this on $M_{23} := N(2) = N(3) = \{1, \dots, 10\} \setminus \{6\}$, and finally variable 4 is always observed, i.e. it is observed on $M_4 := N(4) = \{1, \dots, 10\}$. These three sets form the row ID pattern $\mathcal{N} = \{M_1, M_{23}, M_4\}$. The row partition $I_{\mathcal{N}}$ is given by

$$
\text{(9.12)} \qquad I_{M_1} = \{1\}, \quad I_{M_{23}} = \{2, 3\}, \quad I_{M_4} = \{4\}.
$$

The ordering of the elements in $\mathcal{N}$ is

$$
\text{(9.13)} \qquad M_1 \subsetneq M_{23} \subsetneq M_4,
$$

which implies that $\mathcal{J}(\mathcal{K}_{\mathbb{I}}) = \{4, 234, 1234\}$ and $\mathcal{K}_{\mathbb{I}} = \{\emptyset, 4, 234, 1234\}$. Recall that $\mathcal{K}_{\mathbb{I}}$ could have been generated from the column ID pattern $\mathcal{I} = \{4, 234, 1234\}$. Since $\mathcal{K}_{\mathbb{I}}$ is monotone it imposes no LCIs. Figure 3 illustrates the lattice $\mathcal{K}_{\mathbb{I}}$ and the equivalent TADG $D_{\mathbb{I}}$.

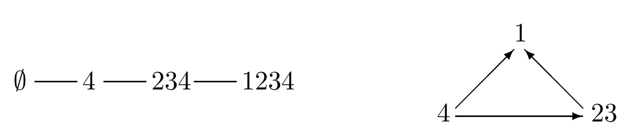

FIGURE 3. The ID lattice $\mathcal{K}_{\mathbb{I}}$ and corresponding TADG $D_{\mathbb{I}}$.

Since $\Gamma_3 Z$ is a MANOVA subspace of $\mathbb{R}^{I\times N}$ we know from Corollary 6.2 that the projected space $\mathcal{U} := p_{\mathbb{I}}(\Gamma_3 Z)$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$ as introduced in Section 6.1. Moreover, we do not face any problems of parameter identifiability (see Remark 6.6). To see this, first note that (6.25) is satisfied, since on at least one subject all variables are observed together, here e.g. all subjects in M_1, yielding $I \in \mathcal{I}$. Second, the $J \times \{1,3,4,7,9\}$ submatrix of Z is of full rank 5 since it is invertible. In detail,

$$Z_{J\times\{1,3,4,7,9\}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \qquad Z^{-1}_{J\times\{1,3,4,7,9\}} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 \end{pmatrix}. \tag{9.14}$$

Hence, since $\{1,3,4,7,9\} \subsetneq M_1, M_{23}, M_4$, the regression spaces for the observed instances of variables 1, 2, 3, and 4 are all 5-dimensional. Thus $\gamma Z \in \Gamma_3 Z$ is identified by $\mu = p_{\mathbb{I}}(\gamma Z) \in \mathcal{U}$.

In conclusion, the LF of the linear ID model $\mathbf{N}(\mathcal{U}) = \mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$ factors as

$$f(1\ 2\ 3\ 4)_{\mathbb{I}} = f(1 \mid 2\ 3\ 4)_{M_1}\ f(2\ 3 \mid 4)_{M_{23}}\ f(4)_{M_4}, \tag{9.15}$$

where the subscripts indicate that the joint distribution is the distribution of the ID array X, that the first factor involves the fully observed block $Y_{1234\times M_1}$, that the second factor involves the fully observed block $Y_{234\times M_{23}}$, and that the third factor involves the fully observed block $Y_{4\times M_4}$. The parameters (μ, Σ) of the model $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$ can be reconstructed from the $\mathcal{K}_{\mathbb{I}}$-parameters defined in (6.18) but condition (6.23) that is necessary and sufficient for the almost sure existence and uniqueness of the MLE of (μ, Σ) is not fulfilled because $|N^{+}_{K_{M_1}}| = |M_1| = 7 \not\geq 9 = 4 + 5 = |K_{M_1}| + \dim(U_1)$, where U_1 is the mean space for the observed instances of variable 1. □

Example 9.3b (*Monotone linear ID model)***.** Assume again the ID pattern from (9.11). Then the ID lattice $\mathcal{K}_{\mathbb{I}}$ and the TADG $D_{\mathbb{I}}$ are the same as in Example 9.3a (see Figure 3). By (2.9), a generalized block-triangular matrix $A \in \mathbf{M}(\mathcal{K}_{\mathbb{I}})$ is of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{pmatrix}. \tag{9.16}$$

Now, let $\Gamma \equiv \tilde{\Gamma}_3$ be the subspace of $\mathbf{M}(I \times J)$ given by the matrices of the form

$$\gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{13} & \gamma_{11} \\ \gamma_{21} & \gamma_{22} & 0 & 0 & \gamma_{21} \\ \gamma_{31} & \gamma_{32} & 0 & 0 & \gamma_{31} \\ \gamma_{41} & 0 & 0 & 0 & \gamma_{41} \end{pmatrix}. \tag{9.17}$$

It is obvious that for $A \in \mathbf{M}(\mathcal{K}_{\mathbb{I}})$ and $\gamma \in \tilde{\Gamma}_3$ the matrix product $A\gamma$ has the zero pattern specified in (9.17) and thus $A\gamma \in \tilde{\Gamma}_3$. This implies that $\mathbf{M}(\mathcal{K}_{\mathbb{I}})\tilde{\Gamma}_3 Z \subseteq \tilde{\Gamma}_3 Z$, hence $\tilde{\Gamma}_3 Z$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I\times N}$. By Corollary 6.2, it follows that $\mathcal{U} := p_{\mathbb{I}}(\tilde{\Gamma}_3 Z)$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$.

It follows that the LF of the linear ID model $\mathbf{N}(\mathcal{U}) = \mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$ factors as in (9.15). By (9.14), there are no parameter identifiability difficulties. Furthermore, with probability one, the MLE of (μ, Σ) exists uniquely because (6.23) holds. To check (6.23) let $\mathcal{U} = U_1 \times (U_{23})^2 \times U_4$ where U_1, U_{23}, and U_4 are the projections of $\mathcal{U}$ onto $\mathbb{R}^{1\times M_1}$, $\mathbb{R}^{23\times M_{23}}$, and $\mathbb{R}^{4\times M_4}$, respectively. Then (6.23) consists of the three inequalities

$$
\begin{aligned}
|N^{+}_{K_{M_1}}| &= |M_1| = 7 \geq 7 = 4 + 3 = |K_{M_1}| + \dim(U_1), \\
|N^{+}_{K_{M_{23}}}| &= |M_{23}| = 9 \geq 5 = 3 + 2 = |K_{M_{23}}| + \dim(U_{23}), \\
|N^{+}_{K_{M_4}}| &= |M_4| = 10 \geq 2 = 1 + 1 = |K_{M_4}| + \dim(U_4),
\end{aligned}
\tag{9.18}
$$

which are fulfilled. □

Example 9.4 (*Nonmonotone linear ID model)*. Consider the subspace $\Gamma \equiv \Gamma_4$ of $\mathbf{M}(I \times J)$ given by the matrices of the form

$$
\gamma = \begin{pmatrix}
\gamma_{11} & \gamma_{11} & \gamma_{13} & \gamma_{14} & \gamma_{11} \\
\gamma_{21} & \gamma_{21} & \gamma_{23} & 0 & \gamma_{21} \\
\gamma_{31} & \gamma_{31} & 0 & \gamma_{34} & \gamma_{31} \\
\gamma_{41} & \gamma_{41} & 0 & 0 & \gamma_{41}
\end{pmatrix}.
\tag{9.19}
$$

Moreover, assume that the ID array has the form

$$
X = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\blacksquare & & \blacksquare & \blacksquare & \blacksquare & & \blacksquare & & \blacksquare & \blacksquare \\
\blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\
\blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & & \blacksquare & & \blacksquare & \blacksquare \\
\blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare
\end{array}.
\tag{9.20}
$$

Then the row ID pattern is $\mathcal{N} = \{M_1, M_2, M_3, M_4\}$ where

$$
\begin{aligned}
&M_1 := N(1) = \{1, 3, 4, 5, 7, 9, 10\}, \qquad M_2 := N(2) = \{1, \ldots, 10\} \setminus \{6\} \\
&M_3 := N(3) = \{1, 2, 3, 4, 5, 7, 9, 10\}, \quad M_4 := N(4) = \{1, \ldots, 10\}.
\end{aligned}
\tag{9.21}
$$

The row partition $I_{\mathcal{N}}$ is given by

$$
I_{M_1} = \{1\}, \quad I_{M_2} = \{2\}, \quad I_{M_3} = \{3\}, \quad I_{M_4} = \{4\}.
\tag{9.22}
$$

From

$$
M_1 \subsetneq M_2, M_3, M_4, \quad M_2 \subsetneq M_4, \quad M_3 \subsetneq M_4,
\tag{9.23}
$$

it follows that $\mathcal{J}(\mathcal{K}_{\mathbb{I}}) = \{4, 24, 34, 1234\}$, $\mathcal{K}_{\mathbb{I}} = \{\emptyset, 4, 24, 34, 234, 1234\}$. Figure 4 shows the graphical representation of $\mathcal{K}_{\mathbb{I}}$ and $D_{\mathbb{I}}$. By (2.9), a generalized block-triangular matrix $A \in \mathbf{M}(\mathcal{K}_{\mathbb{I}})$ is of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & 0 & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{pmatrix}. \tag{9.24}$$

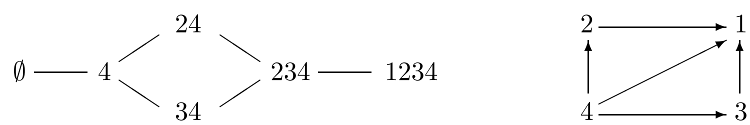

FIGURE 4. The ID lattice $\mathcal{K}_{\mathbb{I}}$ and corresponding TADG $D_{\mathbb{I}}$.

As in Example 9.3b, it is straightforward to verify that $\mathbf{M}(\mathcal{K}_{\mathbb{I}})\Gamma_4 \subseteq \Gamma_4$, which implies that $\Gamma_4 Z$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I\times N}$, which in turn implies by Corollary 6.2 that $\mathcal{U} := p_{\mathbb{I}}(\Gamma_4 Z)$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$.

The LF of the LCI-restricted linear ID model $\mathbf{N}(\mathcal{U}, \mathcal{K}_{\mathbb{I}})$ on $\mathbb{R}^{\mathbb{I}}$, which is based on the CI $2 \perp\!\!\!\perp 3 \mid 4$, factors as

$$f(1\ 2\ 3\ 4)_{\mathbb{I}} = f(1 \mid 2\ 3\ 4)_{M_1}\ f(2 \mid 4)_{M_2}\ f(3 \mid 4)_{M_3}\ f(4)_{M_4}. \tag{9.25}$$

Since $\{1, 3, 4, 7, 9\} \subsetneq M_1, M_2, M_3, M_4$, (9.14) guarantees parameter identifiability. Finally, by (6.23), the MLE of (μ, Σ) exists uniquely almost surely since

$$\begin{aligned} |N^+_{K_{M_1}}| &= |M_1| = 7 \geq 7 = 4 + 3 = |K_{M_1}| + \dim(U_1), \\ |N^+_{K_{M_2}}| &= |M_2| = 9 \geq 4 = 2 + 2 = |K_{M_2}| + \dim(U_2), \\ |N^+_{K_{M_3}}| &= |M_3| = 8 \geq 3 = 2 + 1 = |K_{M_3}| + \dim(U_3), \\ |N^+_{K_{M_4}}| &= |M_4| = 10 \geq 2 = 1 + 1 = |K_{M_4}| + \dim(U_4), \end{aligned} \tag{9.26}$$

where U_i is the projection of $\mathcal{U}$ onto $\mathbb{R}^{i\times M_i}$ for $i = 1, 2, 3, 4$. □

Example 9.5a (*Monotone SUR with monotone incomplete data—no additional CIs)***.**
Let $\Gamma \equiv \Gamma_5$ be the subspace of $\mathbf{M}(I \times J)$ given by the matrices of the form

$$\gamma = \begin{pmatrix} \gamma_{11} & 0 & \gamma_{13} & \gamma_{14} & \gamma_{11} \\ \gamma_{21} & 0 & 0 & 0 & \gamma_{21} \\ \gamma_{31} & 0 & 0 & \gamma_{34} & \gamma_{31} \\ \gamma_{41} & 0 & 0 & 0 & \gamma_{41} \end{pmatrix}, \tag{9.27}$$

and assume incomplete data as in (9.11). The row ID pattern $\mathcal{N}$ is given in Example 9.3a, the associated ID lattice $\mathcal{K}_{\mathbb{I}}$ is shown in Figure 3, and the generalized block-triangular matrices $\mathbf{M}(\mathcal{K}_{\mathbb{I}})$ are of the form (9.16).

Now, choose the particular matrix

$$A^{(0)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbf{M}(\mathcal{K}_{\mathbb{I}}), \tag{9.28}$$

then for $\gamma \in \Gamma_5$ the $(2,4)$ entry $(A^{(0)}\gamma)_{24} = \gamma_{34}$. Thus, $A^{(0)}\gamma \notin \Gamma_5$ if $\gamma_{34} \neq 0$. Therefore, $\Gamma_5 Z$ is not a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I\times N}$. Note, however, that it is still possible that $p_{\mathbb{I}}(\Gamma_5 Z)$ is a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$ (compare the existence statement in Corollary 6.2).

In order to show that $p_{\mathbb{I}}(\Gamma_5 Z)$ is not a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$, consider again $A^{(0)}$ from (9.28) and $\gamma \in \Gamma_5$. It suffices to consider the columns 7 and 10 of X. These columns are fully observed, i.e. $I(7) = I(10) = I = \{1,2,3,4\}$ (compare (5.1)), and thus

$$\begin{aligned} [A^{(0)}p_{\mathbb{I}}(\gamma Z)]_{2\times\{7,10\}} &= [A^{(0)}]_{2\times I}[\gamma Z]_{I\times\{7,10\}} \\ &= [A^{(0)}\gamma]_{2\times J}[Z]_{J\times\{7,10\}} \\ &= \Big(2\gamma_{31} + \gamma_{34}, 2\gamma_{31}\Big). \end{aligned} \tag{9.29}$$

On the other hand, every $\delta \in \Gamma_5$ satisfies

$$\begin{aligned} [p_{\mathbb{I}}(\delta Z)]_{2\times\{7,10\}} &= (\delta Z)_{2\times\{7,10\}} \\ &= \Big(2\delta_{21}, 2\delta_{21}\Big). \end{aligned} \tag{9.30}$$

Therefore, if we choose $\gamma \in \Gamma_5$ with $\gamma_{34} \neq 0$ then

$$[A^{(0)}p_{\mathbb{I}}(\gamma Z)]_{2\times\{7,10\}} \neq [p_{\mathbb{I}}(\delta Z)]_{2\times\{7,10\}} \quad \forall \delta \in \Gamma_5. \tag{9.31}$$

Thus $[A^{(0)}p_{\mathbb{I}}(\gamma Z)] \notin p_{\mathbb{I}}(\Gamma_5 Z)$.

Alternatively, we could have examined the Cartesian product structure of $p_{\mathbb{I}}(\Gamma_5 Z) = \times(U_i \mid i = 1,2,3,4)$, where

$$\begin{array}{lcll} U_1 &=& \operatorname{span}(Z_{1\times M_1} + Z_{5\times M_1}, Z_{3\times M_1}, Z_{4\times M_1}) & \subseteq \mathbb{R}^{M_1}, \\ U_2 &=& \operatorname{span}(Z_{1\times M_{23}} + Z_{5\times M_{23}}) & \subseteq \mathbb{R}^{M_{23}}, \\ U_3 &=& \operatorname{span}(Z_{1\times M_{23}} + Z_{5\times M_{23}}, Z_{4\times M_{23}}) & \subseteq \mathbb{R}^{M_{23}}, \\ U_4 &=& \operatorname{span}(Z_{1\times M_4} + Z_{5\times M_4}) & \subseteq \mathbb{R}^{M_4}. \end{array} \tag{9.32}$$

Since $U_2 \neq U_3$ (which can be verified by an example similar to the one in the previous paragraph), $U_2 \times U_3$ is not a MANOVA subspace of $\mathbb{R}^{23\times M_{23}}$ but $N(2) = N(3) = M_{23}$ and $\{2,3\} = [\{2,3,4\}]$ for the join-irreducible element $\{2,3,4\} \in \mathcal{J}(\mathcal{K}_{\mathbb{I}})$. Therefore, by Proposition 6.1, $p_{\mathbb{I}}(\Gamma_5 Z)$ is not a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$. However, $p_{\mathbb{I}}(\Gamma_5 Z)$ is a SUR/ID subspace of $\mathbb{R}^{\mathbb{I}}$, which can be seen as follows.

The subspace $U_2 \times U_3$ is a SUR subspace induced by the SUR pair $\mathbb{S}_{M_{23}} = (\mathfrak{U}_{M_{23}}, (I_{M_{23}})_{\mathfrak{U}_{M_{23}}})$ for $\mathbb{R}^{I_{M_{23}} \times M_{23}}$, where $\mathfrak{U}_{M_{23}} = \{U_2, U_3\}$ and $(I_{M_{23}})_{\mathfrak{U}_{M_{23}}} = (I_{M_{23},U_2}, I_{M_{23},U_3}) = (\{2\}, \{3\})$. Trivially, for $i = 1$ and $i = 4$, U_i is a SUR subspace induced by the SUR pair $\mathbb{S}_{M_i} = (\mathfrak{U}_{M_i}, (I_{M_i})_{\mathfrak{U}_{M_i}})$ for $\mathbb{R}^{i \times M_i}$, where $\mathfrak{U}_{M_i} = \{U_i\}$ and $(I_{M_i})_{\mathfrak{U}_{M_i}} = (I_{M_i,U_i}) = (\{i\})$. It follows that $\mathcal{U}_{\mathbb{S}} := p_{\mathbb{I}}(\Gamma_5 Z)$ is a SUR/ID subspace of $\mathbb{R}^{\mathbb{I}}$ for the SUR/ID structure for $\mathbb{R}^{\mathbb{I}}$ given by

$$\mathbb{S} := (\mathbb{S}_{M_1}, \mathbb{S}_{M_{23}}, \mathbb{S}_{M_4}). \tag{9.33}$$

Therefore, we can apply the theory developed in Section 7. The set $\mathcal{F}$ defined in (7.7) is

$$\mathcal{F} = \{(M_1, U_1), (M_{23}, U_2), (M_{23}, U_3), (M_4, U_4)\} \tag{9.34}$$

equipped with the partial ordering $\leq_{\mathcal{F}}$ from (7.8), which is specified by

$$\begin{aligned} (M_4, U_4) &\leq_{\mathcal{F}} (M_{23}, U_2), (M_{23}, U_3), (M_1, U_1); \\ (M_{23}, U_2) &\leq_{\mathcal{F}} (M_{23}, U_3), (M_1, U_1); \\ (M_{23}, U_3) &\leq_{\mathcal{F}} (M_1, U_1). \end{aligned} \tag{9.35}$$

The SUR/ID partition $I_{\mathcal{F}}$ is given by

$$I_{M_1,U_1} = \{1\}, \quad I_{M_{23},U_2} = \{2\}, \quad I_{M_{23},U_3} = \{3\}, \quad I_{M_4,U_4} = \{4\}. \tag{9.36}$$

Thus $\mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}) = \{4, 24, 234, 1234\}$ and $\mathcal{K}_{\mathbb{I},\mathbb{S}} = \{\emptyset, 4, 24, 234, 1234\}$. The lattice $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ and the TADG $D_{\mathbb{I},\mathbb{S}}$ are shown in Figure 5. In particular, the lattice $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ is monotone and the TADG $D_{\mathbb{I},\mathbb{S}}$ is complete, so no CIs are imposed. The LF of the SUR/ID model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}) = \mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$ factors as

$$f(1\ 2\ 3\ 4)_{\mathbb{I}} = f(1 \mid 2\ 3\ 4)_{M_1}\ f(3 \mid 2\ 4)_{M_3}\ f(2 \mid 4)_{M_2}\ f(4)_{M_4}. \tag{9.37}$$

No parameter identifiability difficulties arise (compare (9.14)) and the condition (7.17) for almost sure existence and uniqueness of the MLE (μ, Σ) is fulfilled because

$$\begin{aligned} |N^+_{K_{M_1,U_1}}| &= |M_1| = 7 \geq 7 = 4 + 3 = |K_{M_1,U_1}| + \dim(U_1), \\ |N^+_{K_{M_{23},U_2}}| &= |M_{23}| = 9 \geq 3 = 2 + 1 = |K_{M_{23},U_2}| + \dim(U_2), \\ |N^+_{K_{M_{23},U_3}}| &= |M_{23}| = 9 \geq 5 = 3 + 2 = |K_{M_{23},U_3}| + \dim(U_3), \\ |N^+_{K_{M_4,U_4}}| &= |M_4| = 10 \geq 2 = 1 + 1 = |K_{M_4,U_4}| + \dim(U_4). \end{aligned} \tag{9.38}$$

$\emptyset$ — 4 — 24 — 234 — 1234

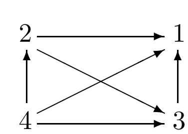

FIGURE 5. The SUR/ID lattice $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ and corresponding TADG $D_{\mathbb{I},\mathbb{S}}$.

Note that even though $\mathcal{U}_{\mathbb{S}}$ is not a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$, the SUR/ID model $\mathbf{N}(\mathcal{U}_{\mathbb{S}})$ can be factored without assuming any LCI restrictions. This is because the lattice $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ is statistically equivalent to the lattice $\mathcal{K}_{\mathbb{I}}$, in the sense that $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I}}) = \mathbf{N}(\mathcal{U}_{\mathbb{S}}) = \mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$. In the ADG context, this subtlety is mentioned in Andersson and Perlman [9, Sect. 12]. □

Example 9.5b (*Monotone SUR with monotone incomplete data—complete independence)***.** Assume incomplete data as in (9.11) with the ID lattice $\mathcal{K}_{\mathbb{I}}$ from Figure 3. Further, suppose that $\Gamma \equiv \Gamma_6$ is the subspace of $\mathbf{M}(I \times J)$ given by the matrices of the form

$$\gamma = \begin{pmatrix} \gamma_{11} & \gamma_{11} & \gamma_{11} & \gamma_{11} & \gamma_{11} \\ \gamma_{21} & \gamma_{22} & \gamma_{22} & \gamma_{22} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{33} & \gamma_{33} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} \end{pmatrix}. \tag{9.39}$$

The mean space $\Gamma_6 Z$ for the unobserved complete random array Y is the Cartesian product of the spaces

$$\begin{aligned} V_1 &= \operatorname{span}(z_1 + z_2 + \cdots + z_5), \quad V_2 = \operatorname{span}(z_1, z_2 + \cdots + z_5), \\ V_3 &= \operatorname{span}(z_1, z_2, z_3 + z_4 + z_5), \quad V_4 = \operatorname{row}(Z). \end{aligned} \tag{9.40}$$

These spaces are totally ordered by inclusion as

$$V_1 \subsetneq V_2 \subsetneq V_3 \subsetneq V_4. \tag{9.41}$$

It follows that $\Gamma_6 Z$ is a SUR subspace induced by the SUR pair $\mathbb{T} = (\mathfrak{U}, I_{\mathfrak{U}})$ for $\mathbb{R}^{I \times N}$, where the SUR pattern $\mathfrak{U} := \{V_1, V_2, V_3, V_4\}$ and the SUR partition $I_{\mathfrak{U}} \equiv (I_{V_i} \mid i = 1, 2, 3, 4) := (\{i\} \mid i = 1, 2, 3, 4)$. Since $\mathfrak{U}$ is monotone the complete data model $\mathbf{N}(\mathcal{U}_{\mathbb{T}}) := \mathbf{N}(\Gamma_6 Z)$ on $\mathbb{R}^{I \times N}$ factors without any CI restrictions.

Similarly, the column or row ID pattern is monotone (compare (9.13)), and the MANOVA model with incomplete data considered in Example 9.3a factors without any CI restrictions (compare the lattice $\mathcal{K}_{\mathbb{I}} = \{\emptyset, 4, 234, 1234\}$ illustrated in Figure 3). If we bring these two aspects together and consider the SUR/ID model on $\mathbb{R}^{\mathbb{I}}$ for the SUR/ID structure $\mathbb{S}$ to be defined in (9.46) then this model factors only under the strongest possible independence assumption $1 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3 \perp\!\!\!\perp 4$, as we now show.

First, $p_{\mathbb{I}}(\Gamma_6 Z)$ is not a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{\mathbb{I}}$ (which implies in particular that $\Gamma_6 Z$ is not a $\mathcal{K}_{\mathbb{I}}$-subspace of $\mathbb{R}^{I \times N}$). This can be seen by considering the matrix

$$A^{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbf{M}(\mathcal{K}_{\mathbb{I}}). \tag{9.42}$$

Then since columns 7 and 10 are fully observed, i.e. $I(7) = I(10) = I = \{1, 2, 3, 4\}$ (compare (5.8)), we obtain similarly as in (9.29) that for $\gamma \in \Gamma_6$

$$[A^{(1)} p_{\mathbb{I}}(\gamma Z)]_{2\times\{7,10\}} = [A^{(1)}\gamma]_{2\times J}[Z]_{J\times\{7,10\}} = \Big(\gamma_{41} + \gamma_{44} + \gamma_{45}, \gamma_{41} + \gamma_{42} + \gamma_{45}\Big). \tag{9.43}$$

Since any $\delta \in \Gamma_6$ satisfies

$$[p_{\mathbb{I}}(\delta Z)]_{2\times\{7,10\}} = (\delta Z)_{2\times\{7,10\}} = \Big(\delta_{21} + 2\delta_{22}, \delta_{21} + 2\delta_{22}\Big), \tag{9.44}$$

then we obtain for $\gamma \in \Gamma_6$ with $\gamma_{42} \neq \gamma_{44}$ that $[A^{(1)} p_{\mathbb{I}}(\gamma Z)] \notin p_{\mathbb{I}}(\Gamma_6 Z)$.

The subspace $p_{\mathbb{I}}(\Gamma_6 Z)$ of $\mathbb{R}^{\mathbb{I}}$ is the Cartesian product $p_{\mathbb{I}}(\Gamma_6 Z) = \times(U_i \mid i = 1, 2, 3, 4)$ with

$$\begin{array}{lcll} U_1 & = & \mathrm{span}(Z_{1\times M_1} + Z_{2\times M_1} + \cdots + Z_{5\times M_1}) & \subseteq \mathbb{R}^{M_1}, \\ U_2 & = & \mathrm{span}(Z_{1\times M_{23}}, Z_{2\times M_{23}} + \cdots + Z_{5\times M_{23}}) & \subseteq \mathbb{R}^{M_{23}}, \\ U_3 & = & \mathrm{span}(Z_{1\times M_{23}}, Z_{2\times M_{23}}, Z_{3\times M_{23}} + Z_{4\times M_{23}} + Z_{5\times M_{23}}) & \subseteq \mathbb{R}^{M_{23}}, \\ U_4 & = & \mathrm{row}(Z_{12345\times M_4}) & \subseteq \mathbb{R}^{M_4}. \end{array} \tag{9.45}$$

Let

$$\mathbb{S} = (\mathbb{S}_{M_1}, \mathbb{S}_{M_{23}}, \mathbb{S}_{M_4}) \tag{9.46}$$

be a SUR/ID structure for $\mathbb{R}^{\mathbb{I}}$, where for $i = 1, 4$, $\mathbb{S}_{M_i} = (\mathfrak{U}_{M_i}, (I_{M_i})_{\mathfrak{U}_{M_i}})$, $\mathfrak{U}_{M_i} = \{U_i\}$, and $(I_{M_i})_{\mathfrak{U}_{M_i}} = (I_{M_i,U_i}) = (\{i\})$. Further, $\mathbb{S}_{M_{23}} = (\mathfrak{U}_{M_{23}}, (I_{M_{23}})_{\mathfrak{U}_{M_{23}}})$, where $\mathfrak{U}_{M_{23}} = \{U_2, U_3\}$ and $(I_{M_{23}})_{\mathfrak{U}_{M_{23}}} = (I_{M_{23},U_2}, I_{M_{23},U_3}) = (\{2\}, \{3\})$. Since $\Gamma_6 Z$ is a SUR subspace of $\mathbb{R}^{I\times N}$ for the SUR pair $\mathbb{T}$, Proposition 7.1 implies that $p_{\mathbb{I}}(\Gamma_6 Z)$ is a SUR/ID subspace of $\mathbb{R}^{\mathbb{I}}$, and in fact $p_{\mathbb{I}}(\Gamma_6 Z) = \mathcal{U}_{\mathbb{S}}$ for the SUR/ID structure $\mathbb{S}$.

The set $\mathcal{F}$ in (7.7) here becomes

$$\mathcal{F} = \{(M_1, U_1), (M_{23}, U_2), (M_{23}, U_3), (M_4, U_4)\}, \tag{9.47}$$

and the SUR/ID partition $I_{\mathcal{F}}$ is given by $I_F = \{i\}$, $i = 1, 2, 3, 4$. The inclusions

$$\begin{array}{lcl} U_1 & \subsetneq & (U_2)_{M_1}, (U_3)_{M_1}, (U_4)_{M_1}, \\ U_2 & \subsetneq & (U_3)_{M_{23}}, (U_4)_{M_{23}}, \\ U_3 & \subsetneq & (U_4)_{M_{23}} \end{array} \tag{9.48}$$

and the inclusions for $\mathcal{N}$ stated in (9.13) imply that no pair of sets in $\mathcal{F}$ is ordered. Thus $\mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}) = \{1, 2, 3, 4\}$ and $\mathcal{K}_{\mathbb{I},\mathbb{S}} = 2^I$, the full power set. The corresponding TADG $D_{\mathbb{I},\mathbb{S}}$ is the empty graph with no edges. In conclusion, the LCI-restricted SUR/ID model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$ is specified by the complete independence $1 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3 \perp\!\!\!\perp 4$, and its LF factors as

$$f(1\ 2\ 3\ 4)_{\mathbb{I}} = f(1)_{M_1}\ f(2)_{M_{23}}\ f(3)_{M_{23}}\ f(4)_{M_4}. \tag{9.49}$$

No parameter identifiability difficulties arise (compare (9.14)) and, by (7.17), the MLE of (μ, Σ) exists uniquely almost surely since

$$\begin{aligned} |N^+_{K_{M_1,U_1}}| &= |M_1| = 7 \geq 2 = 1 + 1 = |K_{M_1,U_1}| + \dim(U_1), \\ |N^+_{K_{M_{23},U_2}}| &= |M_{23}| = 9 \geq 3 = 1 + 2 = |K_{M_{23},U_2}| + \dim(U_2), \\ |N^+_{K_{M_{23},U_3}}| &= |M_{23}| = 9 \geq 4 = 1 + 3 = |K_{M_{23},U_3}| + \dim(U_3), \\ |N^+_{K_{M_4,U_4}}| &= |M_4| = 10 \geq 6 = 1 + 5 = |K_{M_4,U_4}| + \dim(U_4). \end{aligned} \tag{9.50}$$

□

Example 9.6 (*Nonmonotone SUR with nonmonotone incomplete data)***.**
Suppose that the ID array has the form

$$X = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & & & & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & & & & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{pmatrix} \end{array}. \tag{9.51}$$

The row ID pattern is $\mathcal{N} = \{M_1, M_2, M_3, M_4\}$ where

$$\begin{aligned} &M_1 := N(1) = \{1, \ldots, 6, 8, 9, 10\}, \quad M_2 := N(2) = \{1, \ldots, 5, 9, 10\}, \\ &M_3 := N(3) = \{1, 2, 3, 4, 8, 9, 10\}, \quad M_4 := N(4) = \{1, \ldots, 10\}. \end{aligned} \tag{9.52}$$

The row partition $I_{\mathcal{N}}$ is given by $I_{M_i} = \{i\}$ for all $i \in I$. The inclusion ordering of the sets in $\mathcal{N}$ is given by

$$M_1 \subsetneq M_4; \quad M_2 \subsetneq M_1, M_4; \quad M_3 \subsetneq M_1, M_4. \tag{9.53}$$

Now, assume that the mean space is defined through $\Gamma \equiv \Gamma_7$, defined as the subspace of $\mathbf{M}(I \times J)$ given by all matrices of the form

$$\gamma = \begin{pmatrix} 0 & \gamma_{12} & \gamma_{13} & 0 & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & 0 & \gamma_{24} & 0 \\ 0 & \gamma_{32} & 0 & \gamma_{34} & \gamma_{35} \\ 0 & \gamma_{42} & \gamma_{43} & 0 & 0 \end{pmatrix}. \tag{9.54}$$

The complete data subspace $\Gamma_7 Z$ of $\mathbb{R}^{I \times N}$ is the Cartesian product of the four spaces

$$\begin{aligned} &V_1 = \operatorname{span}(z_2, z_3, z_5), \qquad V_2 = \operatorname{span}(z_1, z_2, z_4) \\ &V_3 = \operatorname{span}(z_2, z_4, z_5), \qquad V_4 = \operatorname{span}(z_2, z_3), \end{aligned} \tag{9.55}$$

where $I_{V_i} = \{i\}$. Obviously, $\Gamma_7 Z$ is a SUR subspace of $\mathbb{R}^{I \times N}$ with SUR pair $\mathbb{T} = (\mathfrak{U}, I_{\mathfrak{U}})$ for $\mathbb{R}^{I \times N}$, where the SUR pattern $\mathfrak{U} := \{V_1, V_2, V_3, V_4\}$ and the SUR partition $I_{\mathfrak{U}} \equiv (I_{V_i} \mid i = 1, 2, 3, 4) := (\{i\} \mid i = 1, 2, 3, 4)$.

Let U_i be the projection of V_i onto $\mathbb{R}^{M_i}$. For $i = 1, 2, 3, 4$ define the SUR pairs $\mathbb{S}_{M_i} = (\mathfrak{U}_{M_i}, (I_{M_i})_{\mathfrak{U}_{M_i}})$ for $\mathbb{R}^{\{i\} \times M_i}$ by setting $\mathfrak{U}_{M_i} := \{U_i\}$ and $(I_{M_i})_{\mathfrak{U}_{M_i}} = (I_{M_i, U_i}) := (\{i\})$. Then

$\mathcal{U}_{\mathbb{S}} := p_{\mathbb{I}}(\Gamma_7 Z) = \times(U_i \mid i = 1,2,3,4)$ is a SUR/ID subspace of $\mathbb{R}^{\mathbb{I}}$ for the SUR/ID structure $\mathbb{S} = (\mathbb{S}_{M_i} \mid i = 1,2,3,4)$ for $\mathbb{R}^{\mathbb{I}}$.

The set $\mathcal{F}$ in (7.7) here becomes

$$\mathcal{F} = \{(M_1, U_1), (M_2, U_2), (M_3, U_3), (M_4, U_4)\}. \tag{9.56}$$

In the design matrix specified in (9.10) the two rows z_3 and z_4 are distinct only over the column indices 6 and 7. Those are, however, missing in M_2 and M_3. Therefore, $Z_{3\times M_2} = Z_{4\times M_2}$ and $Z_{3\times M_3} = Z_{4\times M_3}$. This implies that

$$\begin{aligned} (U_4)_{M_1} &\subsetneq U_1 \\ (U_4)_{M_2} &\subsetneq U_2 \\ (U_1)_{M_3}, (U_4)_{M_3} &\subsetneq U_3. \end{aligned} \tag{9.57}$$

By (9.53) and (9.57), the elements of $\mathcal{F}$ are ordered as

$$\begin{aligned} (M_4, U_4) &\leq_{\mathcal{F}} (M_1, U_1), \\ (M_4, U_4) &\leq_{\mathcal{F}} (M_2, U_2), \\ (M_1, U_1), (M_4, U_4) &\leq_{\mathcal{F}} (M_3, U_3). \end{aligned} \tag{9.58}$$

The SUR/ID partition $I_{\mathcal{F}}$ consists of the sets $I_F = \{i\}$ for $F = (M_i, U_i)$. Thus, $\mathcal{J}(\mathcal{K}_{\mathbb{I},\mathbb{S}}) = \{4, 14, 24, 134\}$ and $\mathcal{K}_{\mathbb{I},\mathbb{S}} = \{\emptyset, 4, 14, 24, 124, 134, 1234\}$. Figure 6 shows the graphical representations of $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ and $D_{\mathbb{I},\mathbb{S}}$. The CIs implied by $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ and $D_{\mathbb{I},\mathbb{S}}$ are $2 \perp\!\!\!\perp 13 \mid 4$. The LF of the LCI-restricted SUR/ID model $\mathbf{N}(\mathcal{U}_{\mathbb{S}}, \mathcal{K}_{\mathbb{I},\mathbb{S}})$ on $\mathbb{R}^{\mathbb{I}}$ factors as

$$f(1\,2\,3\,4)_{\mathbb{I}} = f(1 \mid 4)_{M_1}\; f(2 \mid 4)_{M_2}\; f(3 \mid 1\,4)_{M_3}\; f(4)_{M_4}. \tag{9.59}$$

Since $\dim(V_i) = \dim(U_i)$ for all i, no parameter identifiability problems arise. Further, condition (7.17) for the almost sure existence and uniqueness of the MLE of (μ, Σ) is fulfilled since

$$\begin{aligned} |N^{+}_{K_{M_1,U_1}}| &= |M_1| = 9 \geq 5 = 2+3 = |K_{M_1,U_1}| + \dim(U_1), \\ |N^{+}_{K_{M_2,U_2}}| &= |M_3| = 7 \geq 5 = 2+3 = |K_{M_2,U_2}| + \dim(U_2), \\ |N^{+}_{K_{M_3,U_3}}| &= |M_3| = 7 \geq 6 = 3+3 = |K_{M_3,U_3}| + \dim(U_3), \\ |N^{+}_{K_{M_4,U_4}}| &= |M_4| = 10 \geq 3 = 1+2 = |K_{M_4,U_4}| + \dim(U_4). \end{aligned} \tag{9.60}$$

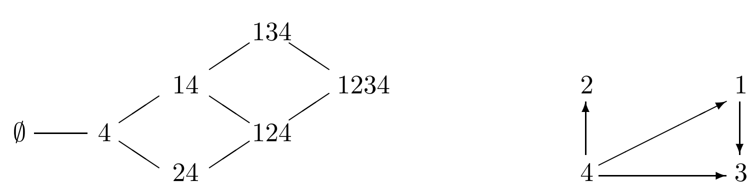

FIGURE 6. The SUR/ID lattice $\mathcal{K}_{\mathbb{I},\mathbb{S}}$ and corresponding TADG $D_{\mathbb{I},\mathbb{S}}$.

Note that in the present example the approach based on the complete data regression spaces as outlined in Remark 7.6 leads to a much more restrictive model based on the lattice $\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}})$. Since the design matrix Z is of full rank, these complete data regression spaces admit only the one inclusion

$$V_4 \subsetneq V_1. \tag{9.61}$$

From (9.61), it is easy to check that $\mathcal{J}(\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}})) = \{2, 3, 4, 14\}$. The graphical representation of $\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}})$ and the equivalent TADG $D(\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}}))$ is given in Figure 7. The CIs imposed by $\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}})$ and $D(\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}}))$ are $14 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3$. □

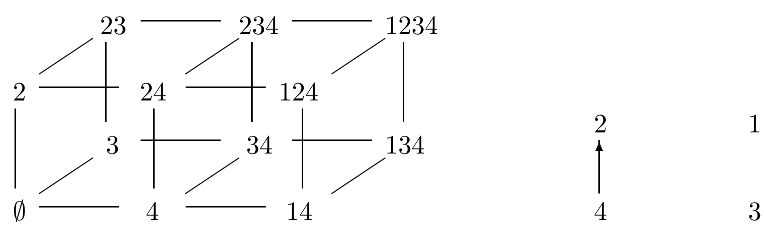

FIGURE 7. The lattice $\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}})$ and corresponding TADG $D(\mathcal{K}(\mathcal{K}_{\mathbb{T}} \cup \mathcal{K}_{\mathbb{I}}))$.

10. Summary and Conclusion

After introducing LCI theory in Sections 2 and 3, we reviewed in Section 4 how the theory can be applied to find minimally restricted SUR models amenable to explicit likelihood inference. In Section 5, we presented the multivariate ID problem. As opposed to standard theory which focuses exclusively on the column ID pattern, we show the importance of the row ID pattern. In Section 6, we extended the work by Andersson and Perlman [5] from i.i.d. incomplete data to linear mean hypotheses which conform to the ID pattern. In particular, we provided a general theory for nonmonotone incomplete data.

Our main results were derived in Section 7 where we found the minimal LCI restrictions that permit us to factor a given SUR/ID model into a product of MANOVA models. In Section 8, we translated the LCI results into the context of graphical Markov models based on TADGs. As can be seen in the examples considered in Section 9, the two equivalent concepts of LCI and TADG models lead to very different graphical representations of CIs. Very small lattices imposing very few CIs can be very easily represented graphically. The equivalent TADG is close to being a complete graph and might be difficult to represent if many variables are observed. On the other hand, it is not possible to give a nice planar illustration of very large lattices imposing many CI constraints. This, however, is easy using a sparse TADG.

Our methods for SUR, for ID problems, and for SUR with incomplete data all rest on the factorization of LCI-restricted models into a product of standard complete data MANOVA models. From the factorization, it is easy to find the explicit MLEs of the regression parameters and to construct the MLEs of the original parameters (see also the reconstruction algorithms in Andersson and Perlman [5, 6, 9]). Moreover, products of MANOVA models have a unimodal LF.

As described in the Introduction, the LCI-restricted model may be used to model the data directly. If this is not desired, the MLE from the LCI-restricted model can be employed to provide new starting values for iterative algorithms such as the EM algorithm or Meng and Rubin's extension to the ECM algorithm [26, 27]. In recent years, there has been much work on speeding up the convergence of the EM algorithm and its extensions (see e.g. Liu and Rubin [23], Liu, Rubin, and Wu [24], and Meng and van Dyk [28]). Our proposed LCI-based starting values require fewer CIs than starting values obtained from ordinary least squares, which assumes complete independence. Since the LCI model is closer to the unrestricted model than the complete independence model, our LCI-based starting values may lead to faster convergence.

Finally, our results are also applicable to testing problems. Suppose that we wish to test a SUR/ID model based on the SUR/ID structure $\mathbb{S}$ for $\mathbb{R}^{\mathbb{I}}$ against a SUR/ID model based on the SUR/ID structure $\tilde{\mathbb{S}}$ for $\mathbb{R}^{\mathbb{I}}$ s.t. $\mathcal{U}_{\mathbb{S}} \subseteq \mathcal{U}_{\tilde{\mathbb{S}}}$ and $\mathbf{N}(\mathcal{U}_{\mathbb{S}}) \subseteq \mathbf{N}(\mathcal{U}_{\tilde{\mathbb{S}}})$. Then the likelihood ratio test statistic can be found explicitly if LCI constraints are imposed which let both models factor into products of MANOVA models. The minimal lattice permitting such a factorization of both models is the lattice generated by the union $\mathcal{K}_{\mathbb{I},\mathbb{S}} \cup \mathcal{K}_{\mathbb{I},\tilde{\mathbb{S}}}$ of the two SUR/ID lattices. The equivalent problem for TADGs is described in Andersson and Perlman [9, Sect.12].

References

1. S.A. Andersson, *The lattice structure of orthogonal linear models and orthogonal variance component models*, Scand. J. Statist. **17** (1990), 287–319.
2. S.A. Andersson, D. Madigan, M.D. Perlman, and C.M. Triggs, *On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs*, Journal of Statistical Planning and Inference **48** (1995), 25–46.
3. ———, *A graphical characterization of lattice conditional independence models*, Annals of Mathematics and Artificial Intelligence **21** (1997), 27–50.
4. S.A. Andersson, J.I. Marden, and M.D. Perlman, *Totally ordered multivariate linear models*, Sankhyā Ser. A **55** (1993), 370–394.
5. S.A. Andersson and M.D. Perlman, *Lattice-ordered conditional independence models for missing data*, Stat. Prob. Letters **12** (1991), 465–486.
6. ———, *Lattice models for conditional independence in a multivariate normal distribution*, Ann. Statist. **21** (1993), 1318–1358.
7. ———, *A characterization of matrix groups that act transitively on the cone of positive definite matrices*, Linear alg. appl. **199** (1994), 151–170.
8. ———, *Normal linear models with lattice conditional independence restrictions*, Multivariate Analysis and its Applications (Hayward, CA), vol. 24, Inst. Math. Statist., 1994, pp. 97–110.
9. ———, *Normal linear regression models with recursive graphical Markov structure*, J. Multivariate Anal. **66** (1998), 133–187.
10. B.A. Davey and H.A. Priestley, *Introduction to Lattices and Order*, second ed., Cambridge University Press, Cambridge, 2002.

11. M. Drton and T.S. Richardson, *Multimodality of the likelihood in the bivariate seemingly unrelated regression model*, Tech. Report 410, Department of Statistics, University of Washington, 2002.
12. D.M. Edwards, *Introduction to Graphical Modelling*, second ed., Springer-Verlag, New York, 2000.
13. A. Goldberger, *A Course in Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1991.
14. G. Graetzer, *General Lattice Theory*, Birkhäuser, Boston, 1978.
15. W.H. Greene, *Econometric Analysis*, third ed., Prentice Hall, Upper Saddle River, New Jersey, 1997.
16. H.-S. Hwang, *Estimation of a linear SUR model with unequal numbers of observations*, The Review of Economics and Statistics **72** (1990), 510–515.
17. J. Kowalski, J.R. Mendoza-Blanco, X.M. Tu, and L.J. Gleser, *On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regressions*, Comm. Statist. Theory Methods **28** (1999), 2119–2140.
18. S.L. Lauritzen, *Graphical Models*, Oxford Statistical Science Series, Clarendon Press, Oxford, UK, 1996.
19. S.L. Lauritzen, A.P. Dawid, B. Larsen, and H.-G. Leimer, *Independence properties of directed Markov fields*, Networks **20** (1990), 491–505.
20. L.P. Lefkovitch, *Seemingly unrelated generalized linear models*, Biometrical J. **33** (1991), 899–912.
21. R.J.A. Little and D.B. Rubin, *Statistical Analysis with Missing Data*, second ed., Wiley, Hoboken, New Jersey, 2002.
22. C.H. Liu, *Efficient ML estimation of the multivariate normal distribution from incomplete data*, J. Multivariate Anal. **69** (1999), 206–217.
23. C.H. Liu and D.B. Rubin, *The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence*, Biometrika **81** (1994), 633–648.
24. C.H. Liu, D.B. Rubin, and Y.N. Wu, *Parameter expansion to accelerate EM: the PX-EM algorithm*, Biometrika **85** (1998), 755–770.
25. K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.
26. X.-L. Meng and D.B Rubin, *Maximum likelihood estimation via the ECM algorithm: A general framework*, Biometrika **80** (1993), 267–278.
27. X.-L. Meng and D.B. Rubin, *Efficient methods for estimation and testing with seemingly unrelated regressions in the presence of latent variables and missing observations*, Bayesian analysis in statistics and econometrics, Wiley, New York, 1996, pp. 215–227.
28. X.L. Meng and D. van Dyk, *The EM algorithm—an old folk-song sung to a fast new tune*, J. Roy. Statist. Soc. Ser. B **59** (1997), 511–567.
29. G.D. Murray, *Comments on "Maximum likelihood from incomplete data via the EM algorithm"*, J. Royal Statist. Soc. Ser. B **39** (1977), 27–28.
30. E.H. Oksanen, *A note on seemingly unrelated regression equations with residual vectors as explanatory variables*, Statist. Probab. Lett. **6** (1987), 103–105.
31. J.L. Schafer, *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London, 1997.
32. P. Schmidt, *Estimation of seemingly unrelated regressions with unequal numbers of observations*, Journal of Econometrics **5** (1977), 365–379.
33. V.K. Sharma, *Estimation of seemingly unrelated regressions with unequal numbers of observations*, Sankhya Ser. B **55** (1993), 135–140.
34. M.A. Steel and G.R. Wood, *On a problem of Andersson and Perlman*, Stat. Prob. Letters **18** (1993), 381–382.
35. P.A.V.B. Swamy and J.S. Mehta, *On Bayesian estimation of seemingly unrelated regressions when some observations are missing*, Journal of Econometrics **3** (1975), 157–169.
36. L.G. Telser, *Iterative estimation of a set of linear regression equations*, J. Amer. Statist. Assoc. **59** (1964), 845–862.
37. K.J. van Garderen, *Curved exponential models in econometrics*, Econometric Theory **13** (1997), 771–790.
38. J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, Wiley, Chichester, 1990.
39. L. Wu and M.D. Perlman, *Lattice conditional independence models for seemingly unrelated regressions*, Comm. Statist. B—Simulation Comput. **29** (2000), 361–384.
40. A. Zellner, *An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias*, J. Amer. Stat. Assoc. **57** (1962), 348–368.
41. ———, *Estimators for seemingly unrelated regression equations: some exact finite sample results*, J. Amer. Stat. Assoc. **58** (1963), 977–992.

Department of Statistics, University of Washington, Seattle, Washington, U.S.A.
E-mail address: drton@stat.washington.edu

Department of Mathematics, Indiana University, Bloomington, Indiana, U.S.A.
E-mail address: standers@indiana.edu

Department of Statistics, University of Washington, Seattle, Washington, U.S.A.
E-mail address: michael@ms.washington.edu