Identifiability of directed Gaussian graphical models with one latent source

Dennis Leung, Mathias Drton

Department of Statistics
University of Washington
Seattle, WA 98195, USA
e-mail: dmhleung@uw.edu; md5@uw.edu

and

Hisayuki Hara

Faculty of Economics
Niigata University
Niigata 950-2181, Japan
e-mail: md5@uw.edu

Abstract: We study parameter identifiability of directed Gaussian graphical models with one latent variable. In the scenario we consider, the latent variable is a confounder that forms a source node of the graph and is a parent to all other nodes, which correspond to the observed variables. We give a graphical condition that is sufficient for the Jacobian matrix of the parametrization map to be full rank, which entails that the parametrization is generically finite-to-one, a fact that is sometimes also referred to as local identifiability. We also derive a graphical condition that is necessary for such identifiability. Finally, we give a condition under which generic parameter identifiability can be determined from identifiability of a model associated with a subgraph. The power of these criteria is assessed via an exhaustive algebraic computational study for small models with 4, 5, and 6 observable variables, and a simulation study for large models with 25 or 35 observable variables.

1. Introduction

In this paper we study parameter identifiability in directed Gaussian graphical models with a latent variable. Our work falls in a line of work where the graphical representation of causally interpretable latent variable models is used to give tractable criteria to decide whether parameters can be uniquely recovered from the joint distribution of the observed variables (Pearl, 2009). Some examples of prior work in this context are Chen et al. (2014), Drton et al. (2011), Foygel et al. (2012), Grzebyk et al. (2004), Kuroki and Miyakawa (2004), Kuroki and Pearl (2014), Stanghellini and Wermuth (2005), Tian (2005), and Tian (2009).

The setup we consider has a single latent variable appear as a source node in the directed graph defining the Gaussian model. The resulting models can be described as follows. Let $X_1, \ldots, X_m$ be observable variables, and let L be a hidden variable, and suppose the variables are related by linear equations as

$$X_v = \sum_{w \neq v} \lambda_{wv} X_w + \delta_v L + \epsilon_v, \qquad v = 1, \ldots, m,$$

where λ_{wv}, δ_v are real coefficients quantifying linear relationships, and the ϵ_v are independent mean zero Gaussian noise terms with variances $\omega_v > 0$. The latent variable L is assumed to be standard normal and independent of the noise terms ϵ_v. Letting $X = (X_1, \ldots, X_m)^T$, $\epsilon = (\epsilon_1, \ldots, \epsilon_m)^T$ and $\delta = (\delta_1, \ldots, \delta_m)^T$, we may present the model in the vectorized form

$$X = \Lambda^T X + \delta L + \epsilon, \tag{1.1}$$

where Λ is the matrix (λ_{wv}) with $\lambda_{vv} = 0$ for all $v = 1, \ldots, m$. We are then interested in specific models, in which for certain pairs of nodes $w \neq v$ the coefficient λ_{wv} is constrained to zero. In particular, we are interested in recursive models, that is, models in which the matrix Λ can be brought into strictly upper triangular form by permuting the indices of the variables (and thus the rows and columns of Λ). This implies that $I_m - \Lambda$ is invertible, where I_m is the $m \times m$ identity matrix. It follows that the observable variate vector X has a m-variate normal distribution $N_m(0, \Sigma)$ with covariance matrix

$$\Sigma = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}, \tag{1.2}$$

where Ω is the diagonal matrix with $\Omega_{vv} = \omega_v$. For additional background on graphical models we refer the reader to Lauritzen (1996) and Pearl (2009). We note that the models we consider also belong to the class of linear structural equation models (Bollen, 1989).

A Gaussian latent variable model postulating recursive zero structure in the matrix Λ from (1.1) can be thought of as associated with a graph $G = (V, E)$ whose vertex set $V = \{1, \ldots, m\}$ is the index set for the observable variables $X_1, \ldots, X_m$. For two distinct nodes $w, v \in V$, the edge set E includes the directed edge (w, v), denoted as $w \to v$ if and only if the model includes λ_{wv} as a free parameter. When the model is recursive, the directed graph G is acyclic and following common terminology we refer to G as a DAG (for directed acyclic graph). In this paper, we will then always assume that the nodes are labeled in topological order, that is, we have $V = \{1, \ldots, m\}$ and $w \to v \in E$ only if $w < v$.

To emphasize the presence of the latent variable L, one could equivalently represent the model by an extended DAG $\overline{G} = (\overline{V}, \overline{E})$ on $m+1$ nodes enumerated as $\overline{V} := \{0, 1, \ldots, m\}$, where the node 0 corresponds to the latent variable L, and if $G = (V, E)$ is the graph on m nodes representing the model as in the preceding paragraph, then $\overline{E} = E \cup \{0 \to v : v \in \{1, \ldots, m\}\}$. The edges $0 \to v$ correspond to the coefficients δ_v.

For the DAG $G = (V, E)$, let

$$\mathbb{R}_E := \{\Lambda = (\lambda_{wv}) \in \mathbb{R}^{m \times m} : w \to v \notin E \Rightarrow \lambda_{wv} = 0\}$$

be the linear space of coefficient matrices, and let diag_m^+ be the set of all $m \times m$ diagonal matrices with a positive diagonal.

Definition 1.1. The *Gaussian one latent source model* associated with a given DAG $G = (V, E)$, denoted as $\mathcal{N}_*(G)$, is the family of all m-variate normal distributions $N_m(0, \Sigma)$ with a covariance matrix of the form

$$\Sigma = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1},$$

for $\Lambda \in \mathbb{R}_E$, $\Omega \in \operatorname{diag}_m^+$ and $\delta \in \mathbb{R}^m$.

The model $\mathcal{N}_*(G)$ has the parametrization map

$$\phi_G \ : \ (\Lambda, \Omega, \delta) \longmapsto (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1} \tag{1.3}$$

defined on the set $\Theta := \mathbb{R}_E \times \operatorname{diag}_m^+ \times \mathbb{R}^m$, which we may also view as an open subset of $\mathbb{R}^{2m+|E|}$, where $|E|$ is the cardinality of the directed edge set E. Clearly, the image of ϕ_G is in PD_m, the cone of positive definite $m \times m$ matrices. Note that since G is acyclic, we have $(I_m - \Lambda)^{-1} = I_m + \Lambda + \Lambda^2 + \cdots + \Lambda^{m-1}$ and thus the covariance parametrization ϕ_G is a polynomial map.

In this paper we will derive graphical conditions that are sufficient/necessary for identifiability of the model $\mathcal{N}_*(G)$. We begin by clarifying what precisely we mean by identifiability. The most stringent notion, namely that of global identifiability, requires ϕ_G to be injective on all of Θ. While this notion is important (Drton et al., 2011), it is too stringent for the setting we consider here. Indeed, for any triple $(\Lambda, \Omega, \delta) \in \Theta$, $\phi_G(\Lambda, \Omega, \delta) = \phi_G(\Lambda, \Omega, -\delta)$, which implies that the *fiber*

$$\{(\Lambda', \Omega', \delta') \in \Theta : \phi_G(\Lambda, \Omega, \delta) = \phi_G(\Lambda', \Omega', \delta')\}$$

always has cardinality ≥ 2. We may account for this symmetry by requiring ϕ_G to be 2-to-1 on all of Θ but this is not enough as there are always some fibers that are infinite. For instance, it is easy to show that the fiber in the above display is infinite when $\delta = 0$. As such, it is natural to consider notions of generic identifiability. Specifically, our contributions will pertain to the notion of generic finite identifiability, as defined below, that only requires finite identification of parameters away from a fixed null set in Θ; here a null set is a set of Lebesgue measure zero. This notion is also referred to as local identifiability in other related work such as Anderson and Rubin (1956).

Null sets appearing in our work are algebraic sets, where an algebraic set $A \subset \mathbb{R}^n$ is the set of common zeros of a collection of multivariate polynomials, i.e.,

$$A = \{a \in \mathbb{R}^n : f_i(a) = 0, \ i = 1, \ldots, k\},$$

for $f_i \in \mathbb{R}[x_1, \ldots, x_n]$, where $\mathbb{R}[x_1, \ldots, x_n]$ is the ring of polynomials in n variables with coefficients in $\mathbb{R}$. If all polynomials f_i are the zero polynomial then $A = \mathbb{R}^n$. Otherwise, A is a proper subset, $A \subsetneq \mathbb{R}^n$, and its dimension is then less than n. In particular, a proper algebraic subset of $\mathbb{R}^n$ has measure zero.

Definition 1.2. Let S be an open subset of $\mathbb{R}^n$, and let f be a map defined on S. Then f is said to be *generically finite-to-one* if there exists a proper algebraic set $\tilde{S} \subset \mathbb{R}^n$ such that the *fiber* of s, i.e. the set $\{s' \in S : f(s') = f(s)\}$, is finite for all $s \in S \setminus \tilde{S}$. Otherwise, f is said to be generically infinite-to-one.

Definition 1.3. The model $\mathcal{N}_*(G)$ of a given DAG $G = (V, E)$ is said to be *generically finitely identifiable* if its parametrization ϕ_G defined on Θ is generically finite-to-one. We also say the DAG G is generically finitely identifiable for short.

Hereafter for any map f defined on an open domain $S \subset \mathbb{R}^n$, we will use

$$\mathcal{F}_f(s) := \{s' \in S : f(s') = f(s)\} \tag{1.4}$$

to denote the fiber of a point $s \in \mathcal{S}$. If T is a subset of S, we will use $f|_T$ to denote the restriction of f to T, in which case for any $t \in T$, we have the fiber

$$\mathcal{F}_{f|_T}(t) = \{t' \in T : f(t') = f(t)\}.$$

The term "generic point" will refer to any point in the domain S that lies outside a fixed proper algebraic subset $\tilde{S}$, and a property is said to hold generically if it holds everywhere on $S \setminus \tilde{S}$. The following well-known lemma is a main tool in this paper, and its proof will be included in Appendix A for completeness. It gives as an immediate corollary a trivial necessary condition for generic finite identifiability.

Lemma 1.1. *Suppose $f : S \to \mathbb{R}^d$ is a polynomial map defined on an open set $S \subset \mathbb{R}^n$. The following statements are equivalent:*

- *(i) f is generically finite-to-one.*
- *(ii) There exists a proper algebraic subset $\tilde{S} \subset \mathbb{R}^n$ such that the fibers of the restricted map $f|_{S \setminus \tilde{S}}$ are all finite, i.e. $|\mathcal{F}_{f|_{S \setminus \tilde{S}}}(s)| < \infty$ for all $s \in S \setminus \tilde{S}$.*
- *(iii) The Jacobian matrix of f is generically of full column rank.*

Corollary 1.2. *Given a DAG $G = (V, E)$, a necessary condition for generic finite identifiability of its associated model $\mathcal{N}_*(G)$ is that $\binom{m+1}{2} - 2m \geq |E|$.*

Proof of . The Jacobian matrix of ϕ_G is of size $\binom{m+1}{2} \times (|E| + 2m)$, and it is necessary that $\binom{m+1}{2} \geq |E| + 2m$ for it to have full column rank. $\square$

Property (ii) is seemingly weaker than (i) in Lemma 1.1. It is useful in proving our results in Section 5. In light of , for the rest of this paper we will restrict our attention to DAGs $G = (V, E)$ with $\binom{m+1}{2} - 2m \geq |E|$, in which case m must be at least 3.

One of our contributions is a sufficient graphical condition stated in Theorem 1.3 below. For $v \neq w \in V$, we will use $v \,\text{---}\, w$ or $w \,\text{---}\, v$ to denote the edge $(v, w) = (w, v)$ of an undirected graph on V. With slight abuse of notation, we may also use $v \,\text{---}\, w$ or $w \,\text{---}\, v$ to denote an edge $v \to w \in E$ when the directionality of edges in a DAG $G = (V, E)$ is to be ignored. For any directed/undirected graph $G = (V, E)$, the complement of G, denoted as $G^c = (V, E^c)$, is the undirected graph on V with the edge set $E^c = \{v \,\text{---}\, w : (v, w) \notin E \text{ and } (w, v) \notin E\}$.
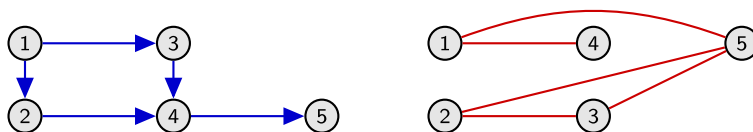
Fɪɢ 1.1. *A DAG G that satisfies the sufficient condition in Theorem 1.3; its undirected complement G^c is shown on the right.*

Theorem 1.3 (Sufficient condition for generic finite identifiability)**.** *The model $\mathcal{N}_*(G)$ given by a DAG $G = (V, E)$ is generically finitely identifiable if every connected component of G^c contains an odd cycle.*

Figure 1.1 shows a DAG G that satisfies the sufficient condition in Theorem 1.3; its undirected complement G^c is shown on the right of the figure. We will revisit this example in Section 4, where we report on algebraic computations that show that for this graph G the fibers of ϕ_G are generically of size 2 or 4.

Our approach to proving Theorem 1.3 also yields a necessary condition for generic finite identifiability. This condition can be stated in terms of two undirected graphs on the node set V, denoted $G_{|L,cov} = (V, E_{|L,cov})$ and $G_{con} = (V, E_{con})$, where $E_{|L,cov}$ captures the dependency of variable pairs after conditioning on the latent variable L, and E_{con} captures the dependency of variable pairs after conditioning on all other variables. From (1.1) it can be seen that $\Sigma_{|L} := (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}$ is the covariance matrix of X conditioning on L, hence $v — w \in E_{|L,cov}$ if and only if $(\Sigma_{|L})_{vw} \not\equiv 0$, and analogously $v — w \in E_{con}$ if and only if $(\Sigma_{|L}^{-1})_{vw} \not\equiv 0$. It is well known that these two undirected graphs can be obtained by using the d-separation criterion applied to the extended DAG $\overline{G}$; see Drton et al. (2009, p. 73) for example.

Theorem 1.4 (Necessary condition for generic finite identifiability)**.** *Given a DAG $G = (V, E)$, for the model $\mathcal{N}_*(G)$ to be generically finitely identifiable, it is necessary that the following two conditions both hold:*

(i) $|E_{con}| - |E| \geq d_{con}$, where d_{con} is the number of connected components in the graph $(G_{con})^c$ that do not contain any odd cycle;
(ii) $|E_{|L,cov}| - |E| \geq d_{cov}$, where d_{cov} is the number of connected components in the graph $(G_{|L,cov})^c$ that do not contain any odd cycle.

Figure 1.2 gives an example of a DAG that fails to satisfy our necessary condition, specifically, condition (*ii*).

In addition to the closely related work of Stanghellini (1997) and Vicard (2000), identifiability of directed Gaussian models with one latent variable has been studied by Stanghellini and Wermuth (2005). The models we treat here are special cases with the latent node being a common parent of all the observable nodes. As we review in more detail in Section 2, we can readily adapt the sufficient graphical criteria given in Stanghellini and Wermuth (2005) for certifying that the model $\mathcal{N}_*(G)$ of a given DAG G is generically finitely identifiable with respect to Definition 1.3. Our own sufficient condition stated in Theorem 1.3
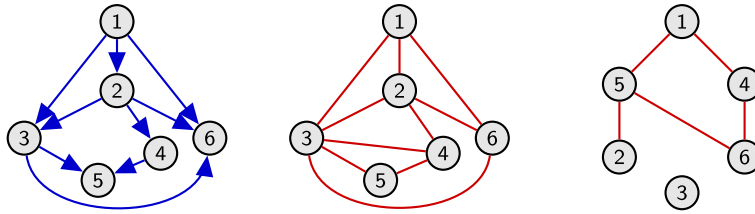
FIG 1.2. *A graph G (left), G_{con} (middle) and G_{con}^c (right). Since $|E_{con}| - |E| = 1 < 2 = d_{con}$, the necessary condition in Thm. 1.4 does not hold.*

is stronger, in the sense that every DAG G satisfying the sufficient conditions in Stanghellini and Wermuth (2005) necessarily satisfies the condition in Theorem 1.3. However, when it applies the result of Stanghellini and Wermuth (2005) yields a stronger conclusion than our generic finiteness result. Indeed as we also emphasize in the discussion in Section 6, their conditions imply that the parmetrization is generically 2-to-1.

We will prove the above stated Theorems 1.3 and 1.4 in Section 3. Since the parametrization map in (1.3) is polynomial, the generic finite identifiability of a given model is decidable by algebraic techniques that involve Gröbner basis computations. In Section 4, we will study the applicability of our graphical criteria via such algebraic computations for all models $\mathcal{N}_*(G)$ of DAGs G with $m = 4, 5, 6$ nodes. Section 4 also contains simulation experiments for larger graphs and a discussion of the computational complexity of checking the graphical conditions. Section 5 will give results on situations where we can determine generic finite identifiability of a model $\mathcal{N}_*(G)$ based on knowledge about the generic finite identifiability of a model $\mathcal{N}_*(G')$, where G' is an induced subgraph of G.

Before ending this introduction, however, we comment on the role that Markov equivalence plays in our problem. Recall that two DAGs defined on the same set of nodes are Markov equivalent if they have the same d-separation relations. The following theorem, which will be proved in Appendix A, says that generic finite identifiability is a property of Markov equivalence classes of DAGs.

Theorem 1.5. *Suppose $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are two Markov equivalent DAGs on the same set of nodes V. Then the model $\mathcal{N}_*(G_1)$ is generically finitely identifiable if and only if the same is true for $\mathcal{N}_*(G_2)$.*

2. Prior work

Stanghellini and Wermuth (2005) give sufficient graphical conditions for identifiability of directed Gaussian graphical models with one latent variable that can be any node in the DAG. We revisit their result in the context of the models from Definition 1.1 and formulate it in terms of generic finite identifiability. (As was mentioned in the Introduction, their result yields in fact the stronger conclusion

of a generically 2-to-1 parametrization.) We begin by stating a well-known fact about DAG models without latent variables.

Lemma 2.1. *For any DAG $G = (V, E)$ with $m = |V|$ nodes, the map*

$$(\Lambda, \Omega) \; \mapsto \; (I_m - \Lambda^T)^{-1} \Omega (I_m - \Lambda)^{-1}$$

is injective on the domain $\mathbb{R}_E \times \mathrm{diag}_m^+$ and has a rational inverse.

Proof. For any $(\Lambda, \Omega) \in \mathbb{R}_E \times \mathrm{diag}_m^+$, let $\Sigma = (\sigma_{vw}) = (I_m - \Lambda^T)^{-1} \Omega (I_m - \Lambda)^{-1}$. Let $pa(v) = \{w : w \to v \in E\}$ be the parent set of the node v. Then one can show, by induction on m and considering a topological ordering of V, that

$$\Lambda_{pa(v),v} \; = \; \left(\Sigma_{pa(v),pa(v)} \right)^{-1} \Sigma_{pa(v),v}$$

and

$$\Omega_{vv} \; = \; \sigma_{vv} - \Sigma_{v,pa(v)} \left(\Sigma_{pa(v),pa(v)} \right)^{-1} \Sigma_{pa(v),v};$$

compare, for instance, Richardson and Spirtes (2002, §8). □

Let the random vector X and the latent variable L have their joint distribution specified via the equation system from (1.1). Write $\Sigma_{|L}$ for the conditional covariance matrix of X given L. Then it holds that

$$\Sigma_{|L} = (I_m - \Lambda^T)^{-1} \Omega (I_m - \Lambda)^{-1}. \tag{2.1}$$

Hence, by Lemma 2.1, when knowing $\Sigma_{|L}$ we can uniquely solve for the pair (Λ, Ω), which are rational functions of $\Sigma_{|L}$. Writing Σ for the (unconditional) covariance matrix of X, we have from (1.2) that

$$\Sigma_{|L} = \Sigma - (I_m - \Lambda^T)^{-1} \delta \delta^T (I_m - \Lambda)^{-1}.$$

Consequently, (Λ, Ω) can be recovered uniquely from Σ and $(I_m - \Lambda^T)^{-1}\delta$. The results of Stanghellini and Wermuth (2005) then address identification of the vector $(I_m - \Lambda^T)^{-1}\delta$, which holds the covariances between each coordinate of X and the latent variable L. We obtain the following observation.

Proposition 2.2 (Adapted from Stanghellini and Wermuth, 2005)**.** *Let $G = (V, E)$ be a DAG. The model $\mathcal{N}_*(G)$ is generically finitely identifiable if*

(i) every connected component of $G_{|L,cov}^c = (V, E_{|L,cov}^c)$ has an odd cycle, or
(ii) every connected component of $G_{con}^c = (V, E_{con}^c)$ has an odd cycle.

Proof. Theorem 1 in Stanghellini and Wermuth (2005) gives (i) or (ii) as a sufficient condition for identifying, up to sign, the m-vector $(I_m - \Lambda^T)^{-1}\delta$ when $\Sigma = \phi_G(\Lambda, \Omega, \delta)$ for a generic point $(\Lambda, \Omega, \delta)$ in Θ. In this case, we can uniquely recover the conditional covariance matrix $\Sigma_{|L}$ from (2.1) and also the pair (Λ, Ω) by Lemma 2.1. After identifying Λ, δ can be solved for, up to sign, by the previous knowledge of $(I_m - \Lambda^T)^{-1}\delta$. Hence, (i) or (ii) is in fact a sufficient condition for generic finite identifiability of $\mathcal{N}_*(G)$. □
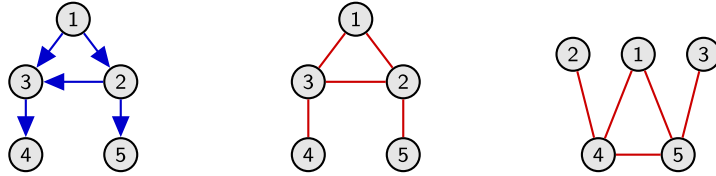
FIG 2.1. *A graph G (left) satisfying the sufficient condition in Proposition 2.2, with G_{con} (middle) and G_{con}^c (right).*

Figure 2.1 shows a DAG with $m = 5$ nodes that satisfies the condition of Proposition 2.2(*ii*).

We conclude this review of prior work by pointing out that any model $\mathcal{N}_*(G)$ that can be determined to be generically finitely identifiable using Proposition 2.2 can also be found to have this property using our new Theorem 1.3.

Proposition 2.3. *A DAG $G = (V, E)$ satisfying either one of the conditions in Proposition 2.2 necessarily satisfies the condition in Theorem 1.3.*

Proof. Let $G_{|L,cov} = (V, E_{|L,cov})$ and $G_{con} = (V, E_{con})$. An edge $v \to w \in E$ also present itself as an undirected edge in both $E_{|L,cov}$ and E_{con}. Hence, when ignoring the directionality of its edges, G is a subgraph of both $G_{|L,cov}$ and G_{con} and, thus, G^c is a supergraph of both $G_{|L,cov}^c$ and G_{con}^c. As such, if every connected component of $G_{|L,cov}^c$, or of G_{con}^c, contains an odd cycle, the same is true of G^c. $\square$

3. Criteria based on the Jacobian of parametrization maps

In this section, we prove Theorems 1.3 and 1.4. Let $G = (V, E)$ be a fixed DAG with $m = |V|$ nodes, and let $\Theta := \mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^m$ denote again the domain of the parametrization

$$\phi_G : (\Lambda, \Omega, \delta) \longmapsto (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}$$

of the covariance matrix of the distributions in model $\mathcal{N}_*(G)$. We begin by introducing other mappings that are generically finite-to-one if and only if ϕ_G is generically finite-to-one.

First, it will be helpful to study the map

$$\tilde{\phi}_G : (\Lambda, \Omega, \delta) \longmapsto (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1} + \delta\delta^T, \qquad (3.1)$$

defined on Θ. Second, focusing on concentration instead of covariance matrices, we will also consider the maps

$$\varphi_G : (\Lambda, \Psi, \gamma) \longmapsto (I_m - \Lambda)(\Psi - \gamma\gamma^T)(I_m - \Lambda^T), \qquad (3.2)$$

$$\tilde{\varphi}_G : (\Lambda, \Psi, \gamma) \longmapsto (I_m - \Lambda)\Psi(I_m - \Lambda^T) - \gamma\gamma^T. \qquad (3.3)$$

Lemma 3.1. *The parametrization ϕ_G is generically finite-to-one if and only if any one of the maps $\tilde{\phi}_G$, φ_G and $\tilde{\varphi}_G$ is generically finite-to-one.*

Proof. Consider first the map $\tilde{\phi}_G$ for which it holds that $\phi_G = \tilde{\phi}_G \circ g$, where

$$g : (\Lambda, \Omega, \delta) \longmapsto (\Lambda, \Omega, (I_m - \Lambda^T)^{-1}\delta)$$

is a diffeomorphism that maps Θ to itself. By the chain rule, the Jacobian of ϕ_G at $(\Lambda, \Omega, \delta)$ is the product of the Jacobian of $\tilde{\phi}_G$ at $g(\Lambda, \Omega, \delta)$ and the Jacobian of g at $(\Lambda, \Omega, \delta)$. Now the latter matrix is invertible on all of Θ since g is a diffeomorphism. It follows that there exists a point in Θ at which the Jacobian of ϕ_G has full column rank if and only if the same is true for $\tilde{\phi}_G$. For the Jacobian of a polynomial map such as ϕ_G and $\tilde{\phi}_G$, full column rank at a single point implies generically full column rank; use the subdeterminants that characterize a drop in rank to define a proper algebraic subset of exceptions, see also Geiger et al. (2001, Lemma 9). The claim about ϕ_G and $\tilde{\phi}_G$ follows from Lemma 1.1.

Let $h : (\Lambda, \Psi, \gamma) \longmapsto (\Lambda, \Psi, (I_m - \Lambda)\gamma)$. Since $\varphi_G = \tilde{\varphi}_G \circ h$, by the same argument as above it also holds that φ_G is generically finite-to-one if and only if $\tilde{\phi}_G$ has this property.

In order to complete the proof of the lemma it suffices to show that ϕ_G is generically finite-to-one if and only if the same holds for φ_G. Define another diffeomorphism from Θ to itself as

$$\rho : (\Lambda, \Omega, \delta) \longmapsto \left(\Lambda, \Omega^{-1}, (1 + \delta^T \Omega^{-1} \delta)^{-1/2} \Omega^{-1} \delta \right).$$

Writing *inv* for matrix inversion, we then have that

$$inv \circ \phi_G = \varphi_G \circ \rho \tag{3.4}$$

because of the identity $(\Omega + \delta \delta^T)^{-1} = (\Psi - \gamma \gamma^T)$ with $\Psi = \Omega^{-1}$ and $\gamma = k^{-1/2} \Psi \delta$, where $k = 1 + \delta^T \Psi \delta > 0$; see e.g. Rao (1973, p. 33). Using (3.4), the equivalence of being generically finite-to-one for ϕ_G and φ_G may be argued similarly as for the maps considered earlier. □

Let $J(\tilde{\varphi}_G)$ be the Jacobian matrix of the map $\tilde{\varphi}_G$ from (3.3). It will be examined to prove Theorem 1.3. In light of Lemmas 1.1 and 3.1, we will show that if G satisfies the condition in Theorem 1.3, then $J(\tilde{\varphi}_G)$ is generically of full column rank, implying that ϕ_G is generically finite-to-one. Our arguments will make use of the following lemma that rests on observations made in Vicard (2000).

Lemma 3.2. *Let $G = (V, E)$ be an undirected graph, and let $f_G : \mathbb{R}^V \to \mathbb{R}^E$ be the map with coordinate functions*

$$f_{G,vw}(x) \;=\; x_v x_w, \quad v - w \in E.$$

Then the Jacobian of f_G has generic rank $m - d$, where $m = |V|$ is the number of nodes and d is the number of connected components of G that do not contain an odd cycle.

Proof. For simpler notation, let $f := f_G$. Let J_f be the Jacobian matrix of the polynomial map f, and let $\ker(J_f)$ be its kernel. By the rank theorem (Rudin, 1976, p. 229), the dimension of $\ker(J_f)$ is generically equal to the dimension of the fiber $\mathcal{F}_f$; recall (1.4). Since $\operatorname{rank}(J_f) = m - \dim(\ker(J_f))$, it suffices to show that $\mathcal{F}_f$ has generic dimension d.

Since the claim is about a generic property, we may restrict the domain of f to the open set $\mathcal{X} := (\mathbb{R} \setminus \{0\})^m$. This assumption is made so that Lemma 1 in Vicard (2000) is applicable later without difficulty. Now, fix a point $y \in f(\mathcal{X}) \subset \mathbb{R}^E$. The elements of the fiber $\mathcal{F}_f(y)$ are the vectors $x \in \mathbb{R}^m$, or equivalently, $x \in \mathcal{X}$, that are solutions to the system of equations

$$y_{vw} = x_v x_w, \quad v - w \in E. \tag{3.5}$$

Let $G_1 = (V_1, E_1), \ldots, G_k = (V_k, E_k)$ be the connected components of G, so that $V_1, \ldots, V_k$ form a partition of V and $E_1, \ldots, E_k$ partition E. Let $k' \leq k$ be the number of connected components containing two nodes at least. Without loss of generality, assume $G_{k'+1}, \ldots, G_k$ are all the connected components with only a single node. Then the equations listed in (3.5) can be arranged to form k' disjoint subsystems indexed by $i = 1, \ldots, k'$. The i-th subsystem has the form

$$y_{vw} = x_v x_w, \quad v - w \in E_i \tag{3.6}$$

and exclusively involves the variables $\{x_v : v \in V_i\}$. By Lemma 1 in Vicard (2000) and also the relevant discussion in the proof of Theorem 1 in the same paper, the solution set to (3.6) either contains two points or can be parametrized by a single free variable in $\mathbb{R}$. The former case arises if and only if G_i contains an odd cycle. It follows that the dimension of the solution set of (3.6) is zero when G_i contains an odd cycle, and it has dimension one if G_i does not contain an odd cycle. In addition, each singleton component $G_i = (V_i, \emptyset)$ for $i = k' + 1, \ldots, k$ provides one additional dimension to the fiber $\mathcal{F}_f(y)$, since the corresponding variables in x are not restricted by any equations. We conclude that the dimension of $\mathcal{F}_f(y)$ equals the number of connected components G_i that do not contain an odd cycle. $\qquad\square$

We return to the object of study, namely, the map $\tilde{\varphi}_G$ which sends the $(2m + |E|)$-dimensional set $\Theta = \mathbb{R}_E \times \operatorname{diag}_m^+ \times \mathbb{R}^m$ to the $\binom{m+1}{2}$-dimensional space of symmetric $m \times m$ matrices. The Jacobian $J(\tilde{\varphi}_G)$ is of size $\binom{m+1}{2} \times (2m + |E|)$, and we index its rows by pairs (v, w) with $1 \leq v < w \leq m$, whereas in Section 1 we assume the vertex set $V = \{1, \ldots, m\}$ to be topologically ordered. We now describe a particular way of arranging the rows and columns of $J(\tilde{\varphi}_G)$.

Define the set of "non-edges" as $N := \{(v, w) : v < w \text{ and } (v, w) \notin E\}$; we will also write $v \not\to w$ to express that $(v, w) \in N$. Also, define $D := \{(v, v) : v \in V\}$, so that $D \cup E \cup N$ index all entries in the upper triangular half of an $m \times m$ symmetric matrix. The rows of $J(\tilde{\varphi}_G)$ are now arranged in the order D, E and N. The columns of $J(\tilde{\varphi}_G)$ are indexed such that partial derivatives with respect to the free input variables in the triple (Λ, Ψ, γ) appear from left to right, in the

order Ψ, Λ and γ. In other words, we partition $J(\tilde{\varphi}_G)$ into 9 blocks as follows:

$$
J(\tilde{\varphi}_G) = \begin{array}{c} \\ D \\ E \\ N \end{array} \begin{bmatrix} \overset{\Psi}{\cdots} & \overset{\Lambda}{\cdots} & \overset{\gamma}{\cdots} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}. \tag{3.7}
$$

The following lemma is obtained by inspection of the partial derivatives of $\tilde{\varphi}_G$. Its proof appears in Appendix A.

Lemma 3.3. *The Jacobian matrix $J(\tilde{\varphi}_G)$ is generically of full column rank provided that the submatrix $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is so.*

We now give the proof of Theorem 1.3.

Proof of Theorem 1.3. By Lemmas 1.1 and 3.3, it suffices to show that the matrix $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is generically of full column rank. For each $v \not\to w \in N$,

$$
[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = \left[(I_m - \Lambda)\Psi(I_m - \Lambda^T)\right]_{vw} - \gamma_v \gamma_w. \tag{3.8}
$$

Note that only the right most term in (3.8) contributes to the partial derivatives of $\tilde{\varphi}_G$ with respect to $\gamma = (\gamma_v)_{v \in \{1,\ldots,m\}}$.

Ignoring the directionality of non-edges in N, define the undirected graph $H = (V, N)$ to which we associate a map f_H as in Lemma 3.2. Then

$$
[J(\tilde{\varphi}_G)]_{N,\gamma} = -J_{f_H}.
$$

But J_{f_H} has generically full column rank by Lemma 3.2 because, in fact, H is equal to the complementary graph G^c for which we assume that all connected components contain an odd cycle. $\qquad\square$

We remark that Theorem 1.3 can also be proven by studying the Jacobian of the map $\tilde{\phi}_G$ from (3.1). We chose to work with $\tilde{\varphi}_G$ above since this allowed us to avoid consideration of the inverse of the matrix $I_m - \Lambda$. For Theorem 1.4, however, we consider both $\tilde{\varphi}_G$ and $\tilde{\phi}_G$.

Proof of Theorem 1.4. We first prove the necessity of condition (i) by showing that if $|E_{con}| - |E| < d_{con}$, then the Jacobian matrix $J(\tilde{\varphi}_G)$ always has row rank less than $2m + |E|$. This implies that it cannot be of full column rank which implies the failure of generic finite identifiability by Lemma 1.1.

As in the proof of Theorem 1.3, we consider the set of non-edges N, which we now partition as $N = N_1 \dot{\cup} N_2$, where $N_1 = \{v \not\to w \in E : v - w \in E_{con}\}$, and $N_2 = N \setminus N_1$. Accordingly, we can partition the submatrix $[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}}$ into two block of rows indexed by N_1 and N_2 as

$$
[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}} = \begin{array}{c} \\ N_1 \\ N_2 \end{array} \begin{bmatrix} \overset{\Psi}{\cdots} & \overset{\Lambda}{\cdots} & \overset{\gamma}{\cdots} \\ 0 & 0 & \cdots \end{bmatrix}. \tag{3.9}
$$

To see that the submatrix $[J(\tilde{\varphi}_G)]_{N_2,\{\Psi,\Lambda\}} = 0$, observe first that an entry of $(I - \Lambda)\Psi(I - \Lambda^T)$ is the zero polynomial if and only if the same is true for $\Sigma_{|L}^{-1}$,

where $\Sigma_{|L}$ is the matrix from (2.1). Second, by definition of E_{con} and N_2, if $(v, w) \in N_2$ then $(\Sigma_{|L}^{-1})_{vw} = 0$.

Next, observe that to prove the necessity of condition (i) it suffices to show that the rank of $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ cannot be larger than $m - d_{con}$. Indeed, if this is true, then there exists a subset $N_2' \subset N_2$ with $|N_2'| = m - d_{con}$, such that the submatrix

$$[J(\tilde{\varphi}_G)]_{\{D,E,N_1,N_2'\},\{\Psi,\Lambda,\gamma\}}$$

has the same rank as the original Jacobian matrix $J(\tilde{\varphi}_G)$. However, the submatrix $[J(\tilde{\varphi}_G)]_{\{D,E,N_1,N_2'\},\{\Psi,\Lambda,\gamma\}}$ has $2m + |E_{con}| - d_{con}$ rows, and thus its rank is less than $2m + |E|$ because under condition (i) we have $|E_{con}| - |E| < d_{con}$. As a result, $J(\tilde{\varphi}_G)$ cannot be of full column rank.

It now remains to show that $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ has rank at most $m - d_{con}$. Observe that the undirected graph (V, N_2) is equal to the complementary graph $(G_{con})^c$. Moreover, $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ is equal to the negative Jacobian of the map $f_{(G_{con})^c}$ that we get by applying the construction from Lemma 3.2 to $(G_{con})^c$; recall the proof of Theorem 1.3. Applying Lemma 3.2, we find that $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ has generic rank $m - d_{con}$, which is also the maximal rank that $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ may have.

The proof of (ii) follows the exact same argument as that of (i), by replacing (a) G_{con} with $G_{|L,cov}$, (b) d_{con} with d_{cov}, (c) $\tilde{\varphi}_G$ with $\tilde{\phi}_G$, (d) Ψ with Ω, (e) γ with δ and (f) $J(\tilde{\varphi}_G)$ with $J(\tilde{\phi}_G)$, where $J(\tilde{\phi}_G)$ is partitioned as

$$J(\tilde{\phi}_G) = \begin{array}{c} \\ D \\ E \\ N \end{array} \begin{array}{ccc} \Omega & \Lambda & \delta \\ \left[\begin{array}{ccc} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{array} \right] \end{array}, \tag{3.10}$$

similarly to (3.7). $\qquad\square$

4. Computations and simulation experiments

According to Theorems 1.3 and 1.4 generic (non-)identifiability can be verified using algorithms that determine the existence of odd cycles for each connected component of a given graph. Efficient algorithms based on depth-first search techniques exist for that purpose: Finding all connected components of a graph $G = (V, E)$, directed or not, requires $O(|V| + |E|)$ operations. If G is connected, determining the oddness of its cycles is equivalent to determining the oddness of the *fundamental cycles* with respect to a spanning tree of G, which requires $O(|V|+|E|+l)$ operations, where l is the sum of the lengths of these fundamental cycles. The relevant algorithms are discussed in Reingold et al. (1977, Chap. 8).

To certify generic finite identifiability of a model $\mathcal{N}_*(G)$ for a DAG G based on Theorem 1.3, its complement G^c has to be submitted as an input to the graph algorithms. When G is *sparse*, one may want to avoid handling the dense complement G^c. The following corollary gives two simple criteria.

Corollary 4.1. *The model $\mathcal{N}_*(G)$ for a DAG $G = (V, E)$ is generically finitely identifiable if*

(i) G *has two connected components, at least one of which is incomplete, or*

(ii) G *has three connected components.*

Proof. We will prove (i) here. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be the two connected components of G such that $V = V_1 \cup V_2$ and $E = E_1 \cup E_2$. Clearly, any two nodes $v \in V_1$ and $w \in V_2$ are adjacent in the complement $G^c = (V, E^c)$. Moreover, two nodes $v, w \in V_1$ are connected in G^c as we may pick any $u \in V_2$ to have $v - u$, $u - w \in E^c$. Similarly, any two nodes in V_2 are connected in G^c. Hence, G^c is a connected graph. The claim is then proven if we can show that G^c contains an odd cycle. By assumption of incompleteness, without loss of generality, there are $v_1, w_1 \in V_1$ such that $v_1 \to w_1, w_1 \to v_1 \notin E_1$. This implies that G^c contains the odd cycle $v_1 - u$, $u - w_1$, $w_1 - v_1$ for arbitrary $u \in V_2$.

The proof of (ii) is similar and simpler, and we omit it. □

4.1. Algebraic computations for small graphs

As explained in Drton (2006, §3) and Garcia-Puente et al. (2010), identifiability properties of a model such as $\mathcal{N}_*(G)$ can be decided using Gröbner basis techniques from computational algebraic geometry (Cox et al., 2007). While these techniques are tractable only for small to moderate size problems, we were able to perform an exhaustive algebraic study of all DAGs $G = (V, E)$ with $m \leq 6$ nodes. Beyond a mere decision on whether the parametrization map ϕ_G is generically 1-to-1, the algebraic methods also provide information about the generic cardinality of the fibers of ϕ_G as a map defined on complex space.

Definition 4.1. For a DAG $G = (V, E)$, let $\phi_G^{\mathbb{C}}$ be the map obtained by extending ϕ_G to the complex domain $\mathbb{C}^{2m+|E|}$. If the (complex) fibers of $\phi_G^{\mathbb{C}}$ are generically of cardinality k, then we say that $\phi_G^{\mathbb{C}}$ is generically k-to-one.

The language of Definition 4.1 allows us to give a refined classification of DAGs G in terms of the identifiability properties of the parametrization of model $\mathcal{N}_*(G)$. Indeed, $\mathcal{N}_*(G)$ is generically finitely identifiable if and only if $\phi_G^{\mathbb{C}}$ is generically k-to-one for some $k < \infty$.

Remark. The generic size of the fibers of $\phi_G^{\mathbb{C}}$ equals the generic size of the fibers of the complex extensions of the three maps from Lemma 3.1. The map $\tilde{\varphi}_G$ has low degree coordinates and tends to be the easiest to work with in algebraic computation. Another approach that can be useful is to adapt the algorithm described in Section 8 of the supplementary material for Foygel et al. (2012). To do this note that for $\Lambda \in \mathbb{C}^E$ there exist complex choices of Ω and δ such that $\phi_G(\Lambda, \Omega, \delta) = \Sigma$ if and only if $(I - \Lambda^T)\Sigma(I - \Lambda)$ is a matrix that is the sum of a diagonal matrix, namely, Ω, and a symmetric matrix of rank 1, namely, $\delta\delta^T$. Whether a matrix is of the latter type can be tested using tetrads, that is, 2×2 subdeterminants involving only off-diagonal entries of the matrix; see also (5.4) below. The tetrads of a matrix form a Gröbner basis (de Loera et al., 1995; Drton et al., 2007).

Table 1 lists out the counts of DAGs $G = (V, E)$, with $4 \leq m \leq 6$ nodes, that have $\phi_G^{\mathbb{C}}$ generically k-to-one, for all possible values of k. The table also gives the

TABLE 1
Counts of unlabeled DAGs G with m nodes, at most $\binom{m+1}{2} - 2m$ edges, and complex parametrization $\phi_G^{\mathbb{C}}$ generically k-to-one. Counts are also given for DAGs that satisfy the sufficient conditions from Thm. 1.3 and Prop. 2.2, and DAGs that fail to satisfy the necessary condition from Thm. 1.4

m	4	5	6
$k < \infty$	5	95	3344
$k = 2$	5	87	2961
$k = 4$	0	8	345
$k = 6$	0	0	24
$k = 8$	0	0	14
Prop. 2.2	5	49	985
Thm. 1.3	5	88	2957
$k = \infty$	1	20	552
Thm. 1.4	1	20	361
Total # of DAGs	6	115	3896

the counts of DAGs satisfying the conditions in Theorems 1.3 and 1.4 as well as Proposition 2.2. DAGs with $\binom{m+1}{2} - 2m < |E|$, which trivially give generically ∞-to-one maps $\phi_G^{\mathbb{C}}$ in view of , are excluded. We emphasize that the counts are with respect to unlabeled DAGs, that is, all DAGs that are isomorphic with respect to relabeling of nodes are counted as one unlabeled graph.

In the considered settings the condition in Theorem 1.3 is very successful in certifying DAGs with a generically finitely identifiable model. For instance, when $m = 6$, it is able to correctly identify 2957 out of 3344 such graphs. The previously known sufficient condition of Stanghellini and Wermuth (2005) identifies 985 of them. Our necessary condition in Theorem 1.4 is also useful in assessing graphs that give generically infinite-to-one models. For instance, when $m = 6$, we find that 361 of 552 such graphs violate the condition; recall the example from Figure 1.2.

While, by Proposition 2.3, our sufficient condition in Theorem 1.3 is stronger than that in Proposition 2.2 for generic finite identifiability, the latter condition, due to Stanghellini and Wermuth (2005), in fact implies that $\phi_G^{\mathbb{C}}$ is generically 2-to-one. For $m = 5$, there are 6 DAGs that satisfy the condition in Theorem 1.3 but give generically 4-to-one maps $\phi_G^{\mathbb{C}}$. The graph from Figure 1.1 is an example. We note that for this DAG G the fibers of $\phi_G^{\mathbb{C}}$ intersect the statistically relevant set Θ in either 2 or 4 points, and both possibilities do occur.

4.2. Simulation study for larger graphs

For DAGs with a large number of nodes m, using exhaustive algebraic computations to determine their identifiability is not feasible. We instead assess the power of our graphical conditions by simulations. For $m = 25$ and $m = 35$, we randomly generate 5000 *labeled* DAGs with k edges, where k ranges from 226 to 275 when $m = 25$ and from 461 to 560 when $m = 35$. Note that for each m, the maximum number of edges in the DAGs we draw equals $\binom{m+1}{2} - 2m$; over that limit the DAGs must give generically non-identifiable models by . The number
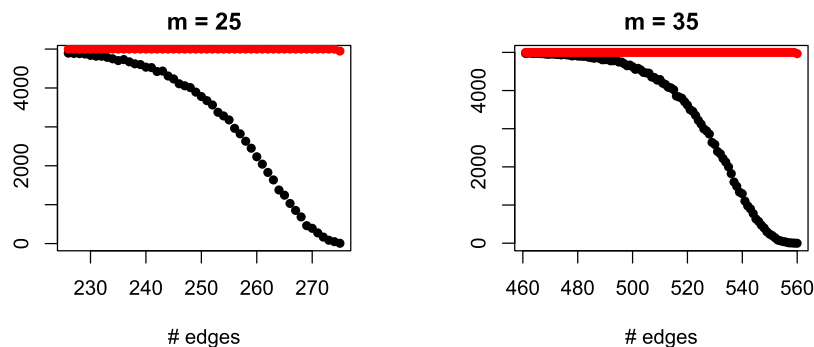
FIG 4.1. *For each $m = 25$ or 35 and each fixed number of edges, the counts of 5000 randomly drawn labeled DAGs satisfying the sufficient condition in Theorem 1.3 are plotted in black, and the counts of labeled DAGs satisfying the necessary condition in Theorem 1.4 are plotted in red.*

of graphs satisfying the conditions in Theorems 1.3 and Theorems 1.4 are plotted in Figure 4.1. As the number of edges increases, our sufficient condition in Theorems 1.3 certifies less of the randomly generated graphs to be identifiable. This agrees with the intuition that it is less likely to have a generically finitely identifiable model as the number of edges increases, since there are more free parameters in the coefficient matrix Λ. On the contrary, the necessary condition in Theorem 1.4 is not very helpful in telling apart generically infinite-to-one models since almost all random graphs satisfy it.

5. Subgraph extension

This section concerns results on how we can extend knowledge about identifiability of an induced subgraph to that of the original DAG. We recall standard terminology in graphical modeling. For a given DAG $G = (V, E)$, we write $pa(v) = \{w : w \to v \in E\}$ for the parent set of the node v, and $ch(v) = \{w : v \to w \in E\}$ for the child set of v. If for some node $s \in V$ there does not exist a node $s' \in V$ with $s \to s' \in E$, then s is a sink node. If there is no other node $s' \in V$ with $s' \to s \in E$, then s is a source node. The following theorem is the main result of this section.

Theorem 5.1. *Given a DAG $G = (V, E)$, if there exists*

   (i) *a sink node $s \in V$ such that $pa(s) \neq V \setminus \{s\}$ and the model $\mathcal{N}_*(G')$ of the induced subgraph G' on $V \setminus \{s\}$ is generically finitely identifiable, or*

   (ii) *a source node $s \in V$ such that $ch(s) \neq V \setminus \{s\}$ and the model $\mathcal{N}_*(G')$ of the induced subgraph G' on $V \setminus \{s\}$ is generically finitely identifiable,*

then the model $\mathcal{N}_*(G)$ is generically finitely identifiable.

Recall that in Table 1 there are $3344 - 2957 = 387$ DAGs with $m = 6$ nodes that are generically finitely identifiable but do not satisfy our sufficient condition

from Theorem 1.3. The above Theorem 5.1 provides a way to certify identifiability of models falling within this "gap", provided that we have knowledge of which DAGs on $m = 5$ nodes are generically finitely identifiable. For instance, from our algebraic computations we know that there are $95 - 88 = 7$ DAGs that are generically finitely identifiable but cannot be proven to be so by Theorem 1.3. Of the 387 aforementioned DAGs on 6 nodes, 194 can be proven to be generically finitely identifiable by using the knowledge about the 7 graphs on $m = 5$ nodes and applying Theorem 5.1. We remark that if a DAG satisfies the condition in Theorem 1.3, the resulting supergraph obtained by augmenting a sink (source) node that does not have every other node as its parent (child) must also satisfy the condition in Theorem 1.3. Hence, given current state-of-the-art, Theorem 5.1 is useful primarily as a tool to reduce the identifiability problem to smaller subgraphs that may then be tackled by algebraic methods.

Theorem 5.1 is obtained by studying the maps ϕ_G and φ_G in (1.3) and (3.2). First consider (1.3). In light of Lemma 1.1(ii), we can show that ϕ_G is generically finite-to-one if there exists a proper algebraic subset $\Xi \subset \mathbb{R}^{2m+|E|}$ such that $|\mathcal{F}_{\phi_G|_{\Theta \setminus \Xi}}(\theta_0)| < \infty$ for all $\theta_0 = (\Lambda_0, \Omega_0, \delta_0) \in \Theta \setminus \Xi$, or equivalently,

$$(I_m - \Lambda^T)\phi_G(\theta_0)(I_m - \Lambda) = \Omega + \delta\delta^T, \tag{5.1}$$

has finitely many solutions for $(\Lambda, \Omega, \delta)$ in $\Theta \setminus \Xi$. *Throughout* this section, Ξ is taken so that all points $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi$ have $\delta_i \neq 0$ for all $i = 1, \ldots, m$. As such, the matrix $\Omega + \delta\delta^T$ on the right hand side of (5.1) has all entries nonzero and is known as a Spearman matrix.

Definition 5.1. A symmetric matrix $\Upsilon \in \mathbb{R}^{m \times m}$ of size $m \geq 3$ is a *Spearman matrix* if $\Upsilon = \Omega + \delta\delta^T$ for a diagonal matrix Ω with positive diagonal and a vector δ with no zero elements.

Any Spearman matrix Υ is positive definite, and it is not difficult to show that if $\Upsilon = \Omega + \delta\delta^T$ is Spearman with $m \geq 3$ then the two summands Ω and $\delta\delta^T$ are uniquely determined as rational functions of Υ. Moreover, $\delta\delta^T$ determines δ up to sign change. For these facts see, for instance, Theorem 5.5 in Anderson and Rubin (1956). We term Ω the *diagonal component* of Υ, and $\delta\delta'$ the *rank-1 component*. The following theorem gives an implicit characterization of Spearman matrices of size $m \geq 4$.

Theorem 5.2. *A positive definite symmetric matrix* $\Upsilon = (\upsilon_{ij}) \in \mathbb{R}^{m \times m}$ *of size $m \geq 4$ is a Spearman matrix if and only if, after sign changes of rows and corresponding columns, all its elements are positive and such that*

$$\upsilon_{ij}\upsilon_{kl} - \upsilon_{ik}\upsilon_{jl} \;=\; \upsilon_{il}\upsilon_{jk} - \upsilon_{ik}\upsilon_{jl} \;=\; \upsilon_{ij}\upsilon_{kl} - \upsilon_{il}\upsilon_{jk} \;=\; 0 \tag{5.2}$$

for $i < j < k < l$, and

$$\upsilon_{ii}\upsilon_{jk} - \upsilon_{ik}\upsilon_{ji} > 0 \tag{5.3}$$

for $i \neq j \neq k$.

This is essentially the same as Theorem 1 in Bekker and de Leeuw (1987), which the reader is referred to for a proof. Unlike Bekker and de Leeuw (1987),

we have a strict inequality in (5.3) since in Definition 5.1 we require the diagonal component of a Spearman matrix to be strictly positive.

The three polynomial expressions in (5.2) are the 2×2 off-diagonal minors of the matrix Υ, which are also known as tetrads in the literature. We call the quadruple $i < j < k < l$ the indices of the tetrad they define. Note that

$$v_{ij}v_{kl} - v_{il}v_{jk} = (v_{ij}v_{kl} - v_{ik}v_{jl}) - (v_{il}v_{jk} - v_{ik}v_{jl})$$

so that the three tetrads in (5.2) are algebraically dependent. In general, a symmetric $m \times m$ matrix Υ has $2\binom{m}{4}$ algebraically independent tetrads and we write TETRADS(Υ) to denote a column vector comprising a choice of $2\binom{m}{4}$ algebraically independent tetrads.

For each triple $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi$ that solves (5.1), it must be true that

$$\text{TETRADS}\left((I_m - \Lambda^T)\phi_G(\theta_0)(I_m - \Lambda)\right) = 0. \tag{5.4}$$

Together with the uniqueness of the diagonal and rank-1 components for a Spearman matrix, if we can show only finitely many Λ's solve the system (5.4), then we have shown that the model $\mathcal{N}_*(G)$ is generically finitely identifiable. Our proof for Theorem 5.1(i) follows this approach.

Alternatively, based on Lemma 3.1, we can also prove generic finite identifiability by considering the map φ_G from (3.2). We then need to show that there exists a proper algebraic subset $\Xi \subset \mathbb{R}^{2m+|E|}$ so that $|\mathcal{F}_{\varphi_G|_{\Theta \setminus \Xi}}(\theta_0)| < \infty$ for all $\theta_0 = (\Lambda_0, \Psi_0, \gamma_0) \in \Theta \setminus \Xi$, or equivalently,

$$(I_m - \Lambda)^{-1}\varphi_G(\theta_0)(I_m - \Lambda^T)^{-1} = \Psi - \gamma\gamma^T \tag{5.5}$$

has finitely many solutions for (Λ, Ψ, γ) in $\Theta \setminus \Xi$. Again we assume that Ξ is defined to avoid issues due to zeros, that is, every triple $(\Lambda, \Psi, \gamma) \in \Theta \setminus \Xi$ has $\gamma_i \neq 0$ for all $i = 1, \dots, m$. We introduce the term coSpearman matrix to describe the matrix on the right hand side of (5.5).

Definition 5.2. A symmetric matrix $\Upsilon \in \mathbb{R}^{m \times m}$ of size $m \geq 3$ is a *coSpearman matrix* if $\Upsilon = \Psi - \gamma\gamma^T$ for a diagonal matrix Ψ with positive diagonal and a vector γ with no zero elements.

Again, the diagonal component Ψ and the rank-1 component $\gamma\gamma^T$ are uniquely determined by Υ; compare Stanghellini (1997, p. 243). The following theorem is analogous to Theorem 5.2.

Theorem 5.3. *A positive definite symmetric matrix $\Upsilon = (v_{ij}) \in \mathbb{R}^{m \times m}$ of size $m \geq 4$ is a coSpearman matrix if and only if, after sign changes of rows and corresponding columns, all its non-diagonal elements are negative and such that*

$$v_{ij}v_{kl} - v_{ik}v_{jl} \;=\; v_{il}v_{jk} - v_{ik}v_{jl} \;=\; v_{ij}v_{kl} - v_{il}v_{jk} \;=\; 0 \tag{5.6}$$

for $i < j < k < l$, and

$$v_{ii}v_{jk} - v_{ik}v_{ji} < 0 \tag{5.7}$$

for $i \neq j \neq k$.

Using the tetrad characterizations (5.6) and the uniqueness of diagonal and rank-1 components, one can now demonstrate that the restricted map $\varphi_G|_{\Theta \setminus \Xi}$ has finite fibers by showing that the system of tetrad equations

$$\text{TETRADS}\left((I_m - \Lambda)^{-1}\varphi_G\left(\theta_0\right)\left(I_m - \Lambda^T\right)^{-1}\right) = 0 \tag{5.8}$$

admits only finitely many solutions for Λ when $\theta_0 \in \Theta \setminus \Xi$.

The finiteness of solutions in Λ for the system (5.4), or (5.8), is a sufficient condition for the generic finite identifiability of $\mathcal{N}_*(G)$. It is, however, not obvious that these two systems necessarily have finitely many solutions when $\mathcal{N}_*(G)$ is generically finitely identifiable. The following lemma states that such a converse does hold for the following two types of DAGs, whose generic finite identifiability can be easily checked by Theorem 1.3. Recall that the notation "$\subsetneq$" means "being a proper subset of".

Lemma 5.4. *Let $G = (V, E)$ be a DAG with vertex set $V = \{1, \ldots, m\}$.*

(i) *If $E \subsetneq \{(k, m) : k \leq m - 1\}$, then there exists a proper algebraic subset Ξ such that for all $\theta_0 = (\Lambda_0, \Omega_0, \delta_0) \in \Theta \setminus \Xi$, the system*

$$\text{TETRADS}\left((I_m - \Lambda^T)\phi_G\left(\theta_0\right)\left(I_m - \Lambda\right)\right) = 0$$

is linear in the variable $\Lambda \in \mathbb{R}_E$ and is solved uniquely by $\Lambda = \Lambda_0$.

(ii) *If $E \subsetneq \{(1, k) : k \geq 2\}$, then there exists a proper algebraic subset Ξ such that for all $\theta_0 = (\Lambda_0, \Psi_0, \gamma_0) \in \Theta \setminus \Xi$, the system*

$$\text{TETRADS}\left((I_m - \Lambda)^{-1}\varphi_G\left(\theta_0\right)\left(I_m - \Lambda^T\right)^{-1}\right) = 0$$

is linear in the variable $\Lambda \in \mathbb{R}_E$ and is solved uniquely by $\Lambda = \Lambda_0$.

The proof of Lemma 5.4 is deferred to Appendix A.

Proof of Theorem 5.1. We will first prove (i), which uses Lemma 5.4(i). The proof of (ii) will follow from similar reasoning using Lemma 5.4(ii).

Without loss of generality, assume that the sink node $s = m$, by giving the nodes a new topological order if necessary. Define two DAGs as follows. First, let $G_1 = (V_1, E_1)$ be the subgraph of G induced by the set $V_1 = V \setminus \{m\} = [m - 1]$, where we adopt the shorthand $[k] := \{1, \ldots, k\}$, $k \in \mathbb{N}$. Second, let $G_2 = (V, E \setminus E_1)$ be the graph on V obtained from G by removing all edges that do not have the sink node m as their head. As before, let $\Theta := \mathbb{R}_E \times \text{diag}_m^+ \times \mathbb{R}^m$. We will construct a proper algebraic subset Ξ, such that for any $\theta \in \Theta \setminus \Xi$, the fiber $\mathcal{F}_{\phi_G|_{\Theta \setminus \Xi}}(\theta)$ is finite. Then Lemma 1.1(ii) applies and yields the assertion of Theorem 5.1(i).

Let $\Theta_1 := \mathbb{R}_{E_1} \times \text{diag}_{m-1}^+ \times \mathbb{R}^{m-1}$, the open set on which the parametrization ϕ_{G_1} of model $\mathcal{N}_*(G_1)$ is defined. By assumption, there exists a proper algebraic subset $\Xi_1' \subset \mathbb{R}^{2(m-1)+|E_1|}$ such that the restricted map $\phi_{G_1}|_{\Theta_1 \setminus \Xi_1'}$ has finite fibers, by Lemma 1.1(ii). Extend Ξ_1' to a proper algebraic subset of $\mathbb{R}^{2m+|E|}$ by defining

$$\Xi_1 := \Xi_1' \times \mathbb{R}^{E \setminus E_1} \times \mathbb{R}^2,$$

where $\mathbb{R}^{E \setminus E_1}$ accommodates the additional free variables λ_{vm} with $v \in pa(m)$, and $\mathbb{R}^2$ accommodates the two variables $\Omega_{mm} = \omega_m$ and δ_m.

Next, recall that for a given point $\theta' = (\Lambda', \Omega', \delta') \in \Theta$, any $(\Lambda, \Omega, \delta) \in \mathcal{F}_{\phi_G}(\theta')$ must satisfy the tetrad equations

$$\text{TETRADS}\left((I_m - \Lambda^T)\phi_G(\theta')(I_m - \Lambda)\right) = 0. \tag{5.9}$$

Let $\lambda_{E_1} := (\lambda_{vw})_{(v,w) \in E_1}^T$. Then any tetrad in (5.9) with indices $i < j < k < m$ has the form

$$\sum_{m' \in pa(m)} a_{m'}\left(\lambda_{E_1}, \phi_G(\theta')\right) \lambda_{m'm} - b\left(\lambda_{E_1}, \phi_G(\theta')\right),$$

where the $a_{m'}$ as well as b are polynomials with the entries of λ_{E_1} and the entries of a symmetric $m \times m$ matrix being their variables. Let $\lambda_{pa(m),m}$ be the vector with entries λ_{vm} for $v \in pa(m)$. Then the part of the system (5.9) involving the variables $\lambda_{v,m}$, $v \in pa(m)$, has the form

$$C\left(\lambda_{E_1}, \phi_G(\theta')\right) \lambda_{pa(m),m} = c\left(\lambda_{E_1}, \phi_G(\theta')\right), \tag{5.10}$$

where C is a matrix of size $2\binom{m-1}{3} \times |pa(m)|$, and c is a vector of length $2\binom{m-1}{3}$. Both C and c are filled with polynomials in the entries of λ_{E_1} and a symmetric $m \times m$ matrix. Since $(\Lambda, \Omega, \delta) \in \mathcal{F}_{\phi_G}(\theta')$, we have $\phi_G(\theta') = \phi_G(\Lambda, \Omega, \delta)$ and, thus,

$$C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right) \lambda_{pa(m),m} = c\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right). \tag{5.11}$$

As θ' was an arbitrary point in Θ, (5.11) holds for all $(\Lambda, \Omega, \delta) \in \Theta$. We claim that $C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right)$ is of full rank for generic choices of $(\Lambda, \Omega, \delta)$. To see this note that if λ_{E_1} is set to 0, then (5.11) becomes the system of tetrad equations for the graph G_2. Using Lemma 5.4(i) and the assumption that $pa(m) \subsetneq V \setminus \{s\}$, we see that $C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right)$ achieves full rank for $\lambda_{E_1} = 0$ and a generic choice of $(\lambda_{pa(m),m}, \Omega, \delta)$. We deduce that the rank is full generically.

Let Ξ_2 be a proper algebraic subset such that $C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right)$ is of full rank for any $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi_2$. Let Ξ_3 be the (algebraic) set comprising all triples $(\Lambda, \Omega, \delta)$ with at least one coordinate $\delta_i = 0$, and define $\Xi := \Xi_1 \cup \Xi_2 \cup \Xi_3$. Clearly, Ξ is a proper algebraic subset of $\mathbb{R}^{2m+|E|}$. Take $(\Lambda_0, \Omega_0, \delta_0)$ to be a point in $\Theta \setminus \Xi$ and define $\Sigma_0 := \phi_G(\Lambda_0, \Omega_0, \delta_0)$. It remains to show that the equation system

$$\Sigma_0 = \phi_G(\Lambda, \Omega, \delta) = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1} \tag{5.12}$$

has only finitely many solutions in $(\Lambda, \Omega, \delta)$ over the set $\Theta \setminus \Xi$.

We begin by observing that because $s = m$ is a sink node, by taking a submatrices in (5.12), we obtain the equation system

$$
\begin{aligned}
(\Sigma_0)_{[m-1]} &= [(I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}]_{[m-1]} \\
&= (I_{m-1} - \Lambda_{[m-1]}^T)^{-1}(\Omega_{[m-1]} + \delta_{[m-1]}\delta_{[m-1]}^T)(I_{m-1} - \Lambda_{[m-1]})^{-1} \\
&= \phi_{G_1}\left(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]}\right).
\end{aligned}
$$

Here, for an index set $W \subset [m]$, we write x_W to denote the subvector $x_W = (x_v : v \in W)$ of vector $x = (x_1, \ldots, x_m)^T$, and we similarly write A_W for the $W \times W$ principal submatrix of a matrix A. Let $\mathcal{S} \subset \Theta_1$ be the projection of the set of all triples $(\Lambda, \Omega, \delta) \in \Theta \setminus \Xi$ that solve (5.12) onto their triple of submatrices/subvector $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]})$. By choice of Ξ, we have that $\mathcal{S} \subset \Theta_1 \setminus \Xi_1'$ and, since $\phi_{G_1}|_{\Theta_1 \setminus \Xi_1'}$ has finite fibers, we know that $\mathcal{S}$ is finite. However, a triple $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]}) \in \mathcal{S}$ determines the matrix C and the vector c in (5.11) and, by choice of Ξ, we may deduce that $\lambda_{pa(m),m}$ is uniquely determined by $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]})$. It follows that the solutions to (5.12) that are in $\Theta \setminus \Xi$ have their Λ part equal to one of $|\mathcal{S}|/2$ many choices; recall that if $(\Lambda_{[m-1]}, \Omega_{[m-1]}, \delta_{[m-1]})$ is in $\mathcal{S}$ then so is $(\Lambda_{[m-1]}, \Omega_{[m-1]}, -\delta_{[m-1]})$. The proof is now complete because Λ determines the Spearman matrix

$$(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda) = \Omega + \delta\delta^T,$$

for which the diagonal component Ω and the rank-1 component $\delta\delta^T$ are uniquely determined. Given the fact that $\delta\delta^T$ determines δ only up to sign, (5.12) has $|\mathcal{S}| < \infty$ solutions over $\Theta \setminus \Xi$, which concludes the proof of (i).

The proof of (ii) is analogous, and we only give a sketch. Instead of considering ϕ_G we turn to φ_G, which also has domain Θ. Without loss of generality, we let the source node be $s = 1$. We then define $G_1 = (V_1, E_1)$ to be the subgraph of G that is induced by $V_1 = \{2, \ldots, m\}$, and we let $G_2 = (V, E \setminus E_1)$. We consider the parametrization φ_{G_1} with domain $\Theta_1 = \mathbb{R}_{E_1} \times \text{diag}_{m-1}^+ \times \mathbb{R}^{m-1}$. By assumption, $\mathcal{N}_*(G_1)$ is generically finitely identifiable, so there exists a proper algebraic subset Ξ_1' such that $\varphi_{G_1}|_{\Theta \setminus \Xi_1'}$ has finite fibers, by Lemma 1.1(ii).

On the other hand, for any $(\Lambda, \Psi, \gamma) \in \Theta$, we have

$$\text{TETRADS}\left((I_m - \Lambda)^{-1}\varphi_G(\Lambda, \Psi, \gamma)(I_m - \Lambda^T)^{-1}\right) = 0.$$

Let $\lambda_{E_1} := (\lambda_{vw})_{(v,w)\in E_1}^T$ and $\lambda_{1,ch(1)} := (\lambda_{1v})_{v \in ch(1)}^T$. Then the tetrad equations with one index equal to $s = 1$ yield the equation system

$$C\left(\lambda_{E_1}, \varphi_G(\Lambda, \Psi, \gamma)\right)\lambda_{1,ch(1)} = c\left(\lambda_{E_1}, \varphi_G(\Lambda, \Psi, \gamma)\right),$$

where part (ii) of Lemma 5.4 can be applied to show that $C\left(\lambda_{E_1}, \varphi_G(\Lambda, \Psi, \gamma)\right)$ is of full rank outside some proper algebraic subset Ξ_2. We may then define a set Ξ as in the proof of part (i) and use arguments similar to the ones above for a proof of part (ii) of our theorem. $\square$

6. Discussion

In this paper we studied identifiability of directed Gaussian graphical models with one latent variable that is a common cause of all observed variables. To our knowledge, the best criteria to decide on identifiability of such models are those given by Stanghellini and Wermuth (2005) who consider a more general setup of Gaussian graphical models with one latent variable. Their results provide a sufficient condition for the strictest notion of identifiability that is meaningful is

this context, namely, whether the parametrization map is generically 2-to-one. Recall that the coefficients associated with the edges pointing from the latent variable to the observables can only be recovered up to a common sign change.

In our work, we take a different approach and study the Jacobian matrix of the parametrization, which leads to graphical criteria to check whether the parametrization is finite-to-one. Our sufficient condition covers all graphs that can be shown to have a 2-to-one parametrization by the conditions of Stanghellini and Wermuth (2005). However, our sufficient condition, which is stated as Theorem 1.3, covers far more graphs as was shown in the computational experiments in Section 4. Our Theorem 1.4 describes a complementary necessary condition.

By studying tetrad equations, we also give a criterion that allows one to deduce identifiability of certain graphs from identifiability of subgraphs (Theorem 5.1). This result is stated for generic finite identifiability but as is clear from the proof, the result would also confirm that the parametrization of a graph is generically 2-to-one provided the involved subgraph has a generically 2-to-one parametrization.

The extension result from Theorem 5.1 can be used in conjunction with the results obtained by the algebraic computations in Section 4. These computations solve the identifiability problem for graphs with up to 6 nodes. In particular, we confirm that the sufficient conditions of Stanghellini and Wermuth (2005) are not necessary for the parametrization map to be generically 2-to-one and provide examples of graphs that yield a generically finite but not 2-to-one parametrization.

As mentioned above, we studied models with one latent source 0 that is connected to all nodes that represent observed variables. However, the graphical criteria in Theorems 1.3 and 1.4 can be readily extended to models with some of these factor loading edges missing. Given the previously used notation, we describe such models as follows. Let $G = (V, E)$ be a DAG with vertex set of size $m = |V|$; these vertices index the observed variables. Let $V' \subset V$ be the nodes representing observed variables that do not directly depend on the latent variable. Then only the edges $0 \to v$ with $v \in V \setminus V'$ are added when forming the extended DAG $\overline{G}$. The parametrization of the Gaussian graphical model determined by G and V' is the restriction of ϕ_G from (1.3) to the domain

$$\Theta(V') := \{(\Lambda, \Omega, \delta) \in \Theta : \delta_v = 0 \text{ for all } v \in V'\}.$$

When the parametrization maps $\tilde{\phi}_G$, φ_G and $\tilde{\varphi}_G$ are restricted to the same domain, the assertion of Lemma 3.1 still holds. The corresponding identifiability results, which are in the spirit of Corollary 1 in Grzebyk et al. (2004), are stated below. A brief outline of their proofs is given in Appendix A.

Theorem 6.1 (Sufficient condition). *Let $G = (V, E)$ be a DAG, and let $V' \subset V$. If every connected component of $(G^c)_{V \setminus V'}$, the subgraph of G^c induced by $V \setminus V'$, contains an odd cycle, then the parametrization map ϕ_G is generically finite-to-one when restricted to the domain $\Theta(V')$.*

The necessary condition given next makes references to the graphs G_{con} and $G_{|L,cov}$ that were defined in the introduction.

Theorem 6.2 (Necessary condition). *Let $G = (V, E)$ be a DAG, and let $V' \subset V$. In order for the restriction of ϕ_G to the domain $\Theta(V')$ to be generically finite-to-one, it is necessary that the following two conditions both hold:*

(i) *Let $\widetilde{G}^c_{con} = (V \setminus V', \widetilde{E}_{con})$ be the subgraph of G^c_{con} induced by $V \setminus V'$. If d_{con} is the number of connected components in the graph $\widetilde{G}^c_{con}$ that do not contain any odd cycle, then $|\widetilde{E}_{con}| - |E| \geq d_{con}$.*

(ii) *Let $\widetilde{G}^c_{|L,cov} = (V \setminus V', \widetilde{E}_{|L,cov})$ be the subgraph of G^c_{con} induced by $V \setminus V'$. If d_{cov} is the number of connected components in the graph $\widetilde{G}^c_{|L,cov}$ that do not contain any odd cycle, then $|\widetilde{E}_{|L,cov}| - |E| \geq d_{cov}$.*

While Theorems 6.1 and 6.2 may be useful in some contexts, models in which latent variables are parents to only some of the observables deserve a more in-depth treatment in future work. In particular, it would be natural to seek ways to combine the results of Stanghellini and Wermuth (2005) and the present paper with the work of Foygel et al. (2012) and Drton and Weihs (2015).

Appendix A: Proofs

Proof of Lemma 1.1. We may assume $d \geq n$, otherwise J_f is never of full column rank. The implication $(i) \Rightarrow (ii)$ is obvious.

To show $(ii) \Rightarrow (iii)$, suppose for contradiction that J_f is not generically of full rank. Since f is polynomial, we then know that $\text{Rank}(J_f) = r < n$ generically, that is, outside a proper algebraic subset $S' \subset \mathbb{R}^n$ the rank is constant r. By the rank theorem (Rudin, 1976, p. 229), for every point $s \in S \setminus (S' \cup \tilde{S})$, we can choose an open ball $\mathcal{B}(s)$ that contains s, is a subset of $S \setminus (S' \cup \tilde{S})$ and for which the restricted map $f|_{\mathcal{B}_s}$ has fibers of dimension $n - r > 0$, contradicting (ii).

It remains to show $(iii) \Rightarrow (i)$. We observe that since f is a polynomial we can assume $S = \mathbb{R}^n$. We then show that the set of points with an infinite fiber, denoted

$$\mathbb{F}_f := \{s \in \mathbb{R}^n : |\mathcal{F}_f(s)| = \infty\},$$

is contained in a proper algebraic subset of $\mathbb{R}^n$. We note that it suffices to assume $n = d$, for without loss of generality, we can permute the d component functions of f and assume that $\pi \circ f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ has a generically full rank Jacobian matrix, where π is the projection onto the first n coordinates. Then $\mathbb{F}_f \subset \mathbb{F}_{\pi \circ f}$.

Now, assume $d = n$, and let $C = \{s \in \mathbb{R}^n : \det J_f(s) = 0\}$ be the set of critical points of f, where J_f is the Jacobian matrix of f. Note that by assumption C is a proper algebraic subset of $\mathbb{R}^n$.

Claim. If $y \in \mathbb{R}^n$ is a point such that $|\mathcal{F}_f(y)| = \infty$, then $\mathcal{F}_f(y) \cap C \neq \emptyset$.

Proof of the Claim. If an algebraic set like $\mathcal{F}_f(y)$ is infinite, then it has dimension $k > 0$. By semialgebraic stratification (Basu et al., 2006), one can see that there exists an open set $U \subset \mathbb{R}^k$ and a differentiable map $g : U \longrightarrow \mathcal{F}_f(y)$ such that the Jacobian of g has full rank on U. If $\mathcal{F}_f(y) \cap C = \emptyset$, then the chain

rule yields that the composition $f \circ g : U \to \{y\}$ has Jacobian of positive rank. This, however, is a contradiction because $f \circ g$ is a constant function. Hence, $\mathcal{F}_f(y) \cap C \neq \emptyset$. □

The claim implies that $\mathbb{F}_f \subset f^{-1}(f(C)) \subset f^{-1}(\overline{f(C)})$, where $\overline{f(C)}$ is the Zariski closure of the semialgebraic set $f(C)$. Since $\overline{f(C)}$ is algebraic, so is $f^{-1}(\overline{f(C)})$ given that f is a polynomial. To finish the proof we only need to show that $f^{-1}(\overline{f(C)})$ has dimension less than n, which is equivalent to $f^{-1}(\overline{f(C)}) \neq \mathbb{R}^n$. By Sard's theorem (Basu et al., 2006, p. 192), $f(C)$, and thus also $\overline{f(C)}$, has dimension less than n. If $f^{-1}(\overline{f(C)}) = \mathbb{R}^n$, then the inverse function theorem, which says that the restricted map $f|_{\mathbb{R}^n \setminus C}$ is a local diffeomorphism, is contradicted. □

Proof of Theorem 1.5. Let $m = |V|$. For $i = 1, 2$, let $\overline{G}_i = (\overline{V}, \overline{E}_i)$ be the extended DAG of G_i, i.e., $\overline{V} = \{0, 1, \ldots, m\}$, and $\overline{E}_i = E_i \cup \{0 \to v : v \in \{1, \ldots, m\}\}$. By the well-known characterization that two DAGs are Markov equivalent if and only if they have the same skeleton and v-structures (Pearl, 2009), it is easy to see that $\overline{G}_1$ and $\overline{G}_2$ are also Markov equivalent.

For $i \in \{1, 2\}$, let $\Theta_i := \mathbb{R}_{E_i} \times \text{diag}_m^+ \times \mathbb{R}^m$. Define

$$\Phi_{\overline{G}_i}\big((\Lambda, \Omega, \delta)\big) = (I_{m+1} - \overline{\Lambda}^T)^{-1}\overline{\Omega}(I_{m+1} - \overline{\Lambda})^{-1},$$

where $\Theta_i := \mathbb{R}_{E_i} \times \text{diag}_m^+ \times \mathbb{R}^m$, $\overline{\Lambda}$ is a $(m+1) \times (m+1)$ matrix such that

$$\overline{\Lambda}_{vw} = \begin{cases} \delta_w & \text{if } v = 0, w = 1, \ldots, m, \\ \Lambda_{vw} & \text{if } v, w = 1, \ldots, m, \\ 0 & \text{otherwise}, \end{cases}$$

and $\overline{\Omega}$ is a diagonal matrix with $\overline{\Omega}_{00} = 1$ and $\overline{\Omega}_{vv} = \Omega_{vv}$ for $v = 1, \ldots m$. Then the image $\Phi_{\overline{G}_i}(\Theta_i)$ is the set of all covariance matrices of $(m+1)$-variate Gaussian distributions that obey the global Markov property of $\overline{G}_i$ and have the variance of node 0, which represents the latent variable L, equal to 1. Consider the projection

$$\pi(\Sigma) = \Sigma_{\{1,\ldots,m\},\{1,\ldots,m\}},$$

where Σ has its rows and columns indexed by $\{0, \ldots, m\}$. Then the parametrization map for the latent variable model $\mathcal{N}_*(G_i)$ equals

$$\phi_{G_i} = \pi \circ \Phi_{\overline{G}_i}. \tag{A.1}$$

Since $\overline{G}_1$ and $\overline{G}_2$ are Markov equivalent, $\Phi_{\overline{G}_1}(\Theta_1) = \Phi_{\overline{G}_2}(\Theta_2)$. By Lemma 2.1, each map $\Phi_{\overline{G}_i}$ is injective on Θ_i with rational inverse defined on the common image $\Phi_{\overline{G}_1}(\Theta_1) = \Phi_{\overline{G}_2}(\Theta_2)$. From (A.1), we obtain that

$$\phi_{G_1} = \pi \circ \Phi_{\overline{G}_1} = \pi \circ \Phi_{\overline{G}_2} \circ \Phi_{\overline{G}_2}^{-1} \circ \Phi_{\overline{G}_1} = \phi_{G_2} \circ \left(\Phi_{\overline{G}_2}^{-1} \circ \Phi_{\overline{G}_1}\right).$$

Since $\Phi_{\overline{G}_2}^{-1} \circ \Phi_{\overline{G}_1} : \Theta_1 \longrightarrow \Theta_2$ is a diffeomorphism, the chain rule implies that the Jacobian of ϕ_{G_1} can be of full column rank if and only if the same is true for ϕ_{G_2}.

Since ϕ_{G_i} are polynomial, the two Jacobians either both have generically full rank or are both everywhere rank deficient. By Lemma 1.1, ϕ_{G_1} is generically finite-to-one if and only if ϕ_{G_2} is so. □

Proof of Lemma 3.3. We first give the structure of $J(\tilde{\varphi}_G)$ block by block.

(a) "$[J(\tilde{\varphi}_G)]_{D,\{\Psi,\Lambda,\gamma\}}$": For a given pair $(v,v) \in D$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vv} = \psi_v + \left(\sum_{w:v \to w \in E} \psi_w \lambda_{vw}^2 \right) - \gamma_v^2.$$

Hence,

$$[J(\tilde{\varphi}_G)]_{(v,v),\psi_w} = \begin{cases} 1 & \text{if } v = w, \\ \lambda_{vw}^2 & \text{if } v \to w \in E, \\ 0 & \text{otherwise,} \end{cases} \tag{A.2}$$

$$[J(\tilde{\varphi}_G)]_{(v,v),\lambda_{wu}} = \begin{cases} 2\lambda_{wu}\psi_u & \text{if } v = w, \\ 0 & \text{otherwise,} \end{cases} \tag{A.3}$$

and

$$[J(\tilde{\varphi}_G)]_{(v,v),\gamma_u} = \begin{cases} -2\gamma_u & \text{if } v = u, \\ 0 & \text{otherwise.} \end{cases} \tag{A.4}$$

(b) "$[J(\tilde{\varphi}_G)]_{E,\{\Psi,\Lambda,\gamma\}}$": For any $v \to w \in E$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = -\lambda_{vw}\psi_w + \left(\sum_{u: \substack{v \to u \in E \\ w \to u \in E}} \lambda_{vu}\lambda_{wu}\psi_u \right) - \gamma_v \gamma_w.$$

Hence,

$$[J(\tilde{\varphi}_G)]_{v \to w, \psi_u} = \begin{cases} -\lambda_{vw} & \text{if } u = w, \\ \lambda_{vu}\lambda_{wu} & \text{if } v \to u \in E \text{ and } w \to u \in E, \\ 0 & \text{otherwise,} \end{cases} \tag{A.5}$$

$$[J(\tilde{\varphi}_G)]_{v \to w, \lambda_{ux}} = \begin{cases} -\psi_w & \text{if } v = u, w = x, \\ \lambda_{wx}\psi_x & \text{if } u = v, u \to x \in E \text{ and } w \to x \in E, \\ \lambda_{vx}\psi_x & \text{if } u = w, u \to x \in E \text{ and } v \to x \in E, \\ 0 & \text{otherwise,} \end{cases} \tag{A.6}$$

and

$$[J(\tilde{\varphi}_G)]_{v \to w, \gamma_u} = \begin{cases} -\gamma_w & \text{if } v = u, \\ -\gamma_v & \text{if } w = u, \\ 0 & \text{otherwise.} \end{cases} \tag{A.7}$$

(c) "$[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}}$": For any $v \not\to w \in N$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = \left(\sum_{u: \substack{v \to u \in E \\ w \to u \in E}} \lambda_{vu} \lambda_{wu} \psi_u \right) - \gamma_v \gamma_w. \qquad (A.8)$$

Hence,

$$[J(\tilde{\varphi}_G)]_{v \not\to w, \psi_u} = \begin{cases} \lambda_{vu}\lambda_{wu} & \text{if } v \to u \in E \text{ and } w \to u \in E, \\ 0 & \text{otherwise}, \end{cases} \qquad (A.9)$$

$$[J(\tilde{\varphi}_G)]_{v \not\to w, \lambda_{ux}} = \begin{cases} \lambda_{wx}\psi_x & \text{if } u = v, \ u \to x \in E \text{ and } w \to x \in E, \\ \lambda_{vx}\psi_x & \text{if } u = w, \ u \to x \in E \text{ and } v \to x \in E, \\ 0 & \text{otherwise}, \end{cases} \qquad (A.10)$$

and

$$[J(\tilde{\varphi}_G)]_{v \not\to w, \gamma_u} = \begin{cases} -\gamma_w & \text{if } v = u, \\ -\gamma_v & \text{if } w = u, \\ 0 & \text{otherwise}. \end{cases} \qquad (A.11)$$

With slight abuse of notation, let $|\Psi|$, $|\gamma|$, $|\Lambda|$ denote the number of free variables in Ψ, γ and Λ respectively. Considering that $|D| = |\Psi|$ and $|E| = |\Lambda|$, we must have that $|N| \geq |\gamma|$ since $J(\tilde{\varphi}_G)$ is a tall matrix. Hence, if $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is generically of full column rank, then there exists a subset $N' \subset N$ such that $|N'| = |\gamma|$ and the determinant of $J(\tilde{\varphi}_G)_{N',\gamma}$ is a nonzero polynomial in the variables of γ, in consideration of (A.11). Now it suffices to show that the $(2m + |E|) \times (2m + |E|)$ square submatrix $[J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}$ is generically of full rank.

Since the concerned matrix has polynomial entries, we need to show that the determinant of $[J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}$ is a nonzero polynomial. To this end, it is sufficient to show that the determinant is a nonzero polynomial in the entries of (Λ, γ) when we specialize $\psi_1 = \cdots = \psi_m = 1$. Noting that $|\Psi| + |\Lambda| + |\gamma| = |D| + |E| + |N'|$, let P denote the set of all permutation functions mapping from the set $D \cup E \cup N'$ to the set of free variables in Λ, Ψ and γ. Choose any ordering of the elements of domain and codomain so as to have a well-defined sign for the permutations. Then by Leibniz's formula, we have

$$\det\left([J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}\right) = \sum_{\sigma \in P} \text{sgn}(\sigma) \prod_{s \in D \cup E \cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)}.$$

Let $\tilde{P}$ be the subset of all permutations $\sigma \in P$ with $\sigma((v,v)) = \psi_v$ for all $(v,v) \in D$ and $\sigma((v,w)) = \lambda_{vw}$ for all $(v,w) \in E$. Then we obtain that

$$\det\left([J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}\right)$$
$$= \sum_{\sigma \in \tilde{P}} \text{sgn}(\sigma) \prod_{s \in D \cup E \cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)} + \sum_{\sigma \in P \setminus \tilde{P}} \text{sgn}(\sigma) \prod_{s \in D \cup E \cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)}$$

$$= \pm \det\big(J(\tilde{\varphi}_G)_{N',\gamma}\big) + \sum_{\sigma \in P \setminus \tilde{P}} \operatorname{sgn}(\sigma) \prod_{s \in D \cup E \cup N'} J(\tilde{\varphi}_G)_{s,\sigma(s)}, \tag{A.12}$$

where the equality in (A.12) follows from (A.2), (A.6) and the fact that $\psi_1 = \cdots = \psi_m = 1$. We also deduce from (A.2)–(A.11) that every summand in the second term of (A.12) is either zero or a polynomial term involving free variables of Λ. In contrast, $\det\big(J(\tilde{\varphi}_G)_{N',\gamma}\big)$ is a nonzero polynomial only in free variables of γ and can thus not be canceled by the second term in (A.12). $\qquad\square$

Proof for Lemma 5.4. We first prove (i). Since $\mathcal{N}_*(G)$ is generically finitely identifiable by Theorem 1.3, there exists an algebraic subset Ξ' such that for all $\theta \in \Theta \setminus \Xi'$, $|\mathcal{F}_{\phi_G}(\theta)| < \infty$. Define Ξ to be the union of Ξ' and the set of triples $(\Lambda, \Omega, \delta) \in \mathbb{R}^{2m+|E|}$ with at least one coordinate $\delta_i = 0$. Let $\Sigma_0 = \phi_G(\Lambda_0, \Omega_0, \delta_0)$ and

$$S = (s_{ij}) := (I_m - \Lambda^T)\Sigma_0(I_m - \Lambda). \tag{A.13}$$

Then for $1 \le i < j \le m$,

$$s_{ij} = \sum_{1 \le k, k' \le m} \lambda_{ki}[\Sigma_0]_{kk'}\lambda_{k'j} - \sum_{1 \le k \le m} [\Sigma_0]_{ik}\lambda_{kj} - \sum_{1 \le k \le m} \lambda_{ki}[\Sigma_0]_{kj} + [\Sigma_0]_{ij}$$

$$= \begin{cases} -\sum_{(k,m) \in E} [\Sigma_0]_{ik}\lambda_{km} + [\Sigma_0]_{im} & \text{if } j = m, \\ [\Sigma_0]_{ij} & \text{if } j < m, \end{cases},$$

where the last equality follows from the fact that λ_{ij} are nonzero only when $(i, j) \in E$. Hence, for any four indices $1 \le i < j < k < l \le m$, the tetrads

$$s_{ij}s_{kl} - s_{ik}s_{jl}, \qquad\qquad s_{il}s_{jk} - s_{ik}s_{jl}$$

are constant polynomials when $l < m$ and have degree 1 in the variables $\{\lambda_{vm} : (v, m) \in E\}$ when $l = m$. The equation system

$$\text{TETRADS}(S) = 0$$

is a thus a consistent linear system that can be represented as

$$C\lambda_{pa(m),m} = c, \tag{A.14}$$

where $\lambda_{pa(m),m} = (\lambda_{vm})^T_{v \in pa(m)}$ is the vector of all free Λ variables, C is a $2\binom{m-1}{3} \times |pa(m)|$ matrix and c is a $2\binom{m-1}{3}$-vector. Both C and c depend only on Σ_0.

To finish the proof, we now need to show that (A.14) is uniquely solvable in $\lambda_{pa(m),m}$. We will aim to contradict $|\mathcal{F}_{\phi_G}(\theta_0)| < \infty$ if (A.14) does not have a unique solution. Note that the solution set is an affine subspace $\mathcal{L} \subset \mathbb{R}^{|E|}$. For a contradiction, suppose that $\mathcal{L}$ is of positive dimension. Upon substituting $\Lambda = \Lambda_0$ into (A.13), we obtain

$$S_0 = (s^0_{ij}) = (I_m - \Lambda_0^T)\Sigma_0(I_m - \Lambda_0),$$

and in consideration of (5.3) in Theorem 5.2, it must be true that

$$s_{ii}^0 s_{jk}^0 - s_{ik}^0 s_{ji}^0 > 0, \text{ for all } i \neq j \neq k.$$

We may then pick an open ball $\mathcal{B}(\Lambda_0)$ such that for all solutions $\Lambda \in \mathcal{L} \cap \mathcal{B}(\Lambda_0)$, the matrix $S = (s_{ij})$ defined by (A.13) satisfies

$$s_{ii} s_{jk} - s_{ik} s_{ji} > 0 \text{ , for all } i \neq j \neq k.$$

It follows that $\mathcal{L} \cap \mathcal{B}(\Lambda_0)$ is an infinite set whose elements Λ all make the matrix $(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda)$ a Spearman matrix. Hence, the system

$$(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda) = \Omega + \delta\delta'$$

has infinitely many solutions, contradicting $|\mathcal{F}_{\phi_G}(\theta_0)| < \infty$.

The proof of (ii) is analogous. We first let $\Upsilon_0 = \varphi_G(\Lambda_0, \Psi_0, \gamma_0)$ and define

$$\tilde{S} = (\tilde{s}_{ij}) \ = \ (I_m - \Lambda)^{-1}\Upsilon_0(I_m - \Lambda^T)^{-1}. \tag{A.15}$$

Noting that in this case $(I_m - \Lambda)^{-1} = I_m + \Lambda$, it can be easily seen that

$$\text{TETRADS}(\tilde{S}) = \text{TETRADS}\big((I_m - \Lambda)^{-1}\Upsilon_0(I_m - \Lambda^T)^{-1}\big) = 0$$

is a linear system in the variables $\{\lambda_{1v} : v \in ch(1)\}$. Similar to the above arguments, we may use Theorem 1.3 and Theorem 5.3 to prove by contradiction that the system can only have a unique solution in $\{\lambda_{1v} : v \in ch(1)\}$. $\qquad\square$

Proof of Theorems 6.1 and 6.2. For Theorem 6.1, one can partition the Jacobian matrix $J(\tilde{\varphi}_G)$ of $\tilde{\varphi}_G$ as in (3.7), only with γ replaced by $\gamma_{V \setminus V'} = \{\gamma_v : v \in V \setminus V'\}$. In analogy with Lemma 3.3, it can be shown that $J_{\tilde{\varphi}_G}$ is of column full rank if $[J(\tilde{\varphi}_G)]_{N, \gamma_{V \setminus V'}}$ is. The reasoning is then analogous to that in the proof of Theorem 1.3, the main step being the application of Lemma 3.2 where the graph defining the considered map becomes $(G^c)_{V \setminus V'}$.

The proof of Theorem 6.2 is analogous to the proof of Theorem 1.4. The only change is to replace G_{con}^c, $G_{|L,cov}^c$, γ and δ by $\tilde{G}_{con}^c$, $\tilde{G}_{|L,cov}^c$, $\gamma_{V \setminus V'}$ and $\delta_{V \setminus V'}$, respectively. $\qquad\square$

Acknowledgments

References

ANDERSON, T. W. and RUBIN, H. (1956). "Statistical inference in factor analysis." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. V*, 111–150. University of California Press, Berkeley and Los Angeles. MR0084943

BASU, S., POLLACK, R., and ROY, M.-F. (2006). *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, second edition. MR2248869

BEKKER, P. A. and DE LEEUW, J. (1987). "The rank of reduced dispersion matrices." *Psychometrika*, 52(1): 125–135. MR0903651

BOLLEN, K. A. (1989). *Structural equations with latent variables*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication. MR0996025

CHEN, B., TIAN, J., and PEARL, J. (2014). "Testable Implications of Linear Structural Equations Models." In BRODLEY, C. E. and STONE, P. (eds.), *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2424–2430. AAAI Press.

COX, D., LITTLE, J., and O'SHEA, D. (2007). *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer, New York, third edition. An introduction to computational algebraic geometry and commutative algebra. MR2290010

DE LOERA, J. A., STURMFELS, B., and THOMAS, R. R. (1995). "Gröbner bases and triangulations of the second hypersimplex." *Combinatorica*, 15(3): 409–424. MR1357285

DRTON, M. (2006). "Algebraic techniques for Gaussian models." In HUŠKOVÁ, M. and JANŽURA, M. (eds.), *Prague Stochastics*, 81–90. Charles University Prague: Matfyzpress.

DRTON, M., FOYGEL, R., and SULLIVANT, S. (2011). "Global identifiability of linear structural equation models." *Ann. Statist.*, 39(2): 865–886. MR2816341

DRTON, M., STURMFELS, B., and SULLIVANT, S. (2007). "Algebraic factor analysis: tetrads, pentads and beyond." *Probab. Theory Related Fields*, 138(3-4): 463–493. MR2299716

DRTON, M., STURMFELS, B., and SULLIVANT, S. (2009). *Lectures on algebraic statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel. MR2723140

DRTON, M. and WEIHS, L. (2015). "Generic identifiability of linear structural equation models by ancestor decomposition." *ArXiv e-prints*. 1504.02992.

FOYGEL, R., DRAISMA, J., and DRTON, M. (2012). "Half-trek criterion for generic identifiability of linear structural equation models." *Ann. Statist.*, 40(3): 1682–1713. MR3015040

GARCIA-PUENTE, L. D., SPIELVOGEL, S., and SULLIVANT, S. (2010). "Identifying causal effects with computer algebra." In GRÜNWALD, P. and SPIRTES, P. (eds.), *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI)*. AUAI Press.

GEIGER, D., HECKERMAN, D., KING, H., and MEEK, C. (2001). "Stratified exponential families: graphical models and model selection." *Ann. Statist.*, 29(2): 505–529. MR1863967

GRZEBYK, M., WILD, P., and CHOUANIÈRE, D. (2004). "On identification of multi-factor models with correlated residuals." *Biometrika*, 91(1): 141–151. MR2050465

KUROKI, M. and MIYAKAWA, M. (2004). "Graphical identifiability criteria for total effects in studies with an unobserved response variable." *Behaviormetrika*, 31(1): 13–28. MR2037887

KUROKI, M. and PEARL, J. (2014). "Measurement bias and effect restoration in causal inference." *Biometrika*, 101(2): 423–437. MR3215357

LAURITZEN, S. L. (1996). *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York. Oxford Science Publications. MR1419991

PEARL, J. (2009). *Causality*. Cambridge University Press, Cambridge, second edition. Models, reasoning, and inference. MR2548166

RAO, C. R. (1973). *Linear statistical inference and its applications*. John Wiley & Sons, New York-London-Sydney, second edition. Wiley Series in Probability and Mathematical Statistics. MR0346957

REINGOLD, E. M., NIEVERGELT, J., and DEO, N. (1977). *Combinatorial algorithms: theory and practice*. Prentice-Hall, Inc., Englewood Cliffs, N.J. MR0471431

RICHARDSON, T. and SPIRTES, P. (2002). "Ancestral graph Markov models." *Ann. Statist.*, 30(4): 962–1030. MR1926166

RUDIN, W. (1976). *Principles of mathematical analysis*. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third edition. International Series in Pure and Applied Mathematics. MR0385023

STANGHELLINI, E. (1997). "Identification of a single-factor model using graphical Gaussian rules." *Biometrika*, 84(1): 241–244.

STANGHELLINI, E. and WERMUTH, N. (2005). "On the identification of path analysis models with one hidden variable." *Biometrika*, 92(2): 337–350. MR2201363

TIAN, J. (2005). "Identifying direct causal effects in linear models." In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 346–352. AAAI Press/The MIT Press.

TIAN, J. (2009). "Parameter identification in a class of linear structural equation models." In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1970–1975. AAAI Press.

VICARD, P. (2000). "On the identification of a single-factor model with correlated residuals." *Biometrika*, 87(1): 199–205. MR1766840