Identifying causal effects with computer algebra

Luis David García–Puente

Department of Mathematics and Statistics
Sam Houston State University University

Parameter Identification In Graphical Models
The American Institute of Mathematics
October 4, 2010

1. The relationships among a set of observed variables are expressed by **linear equations**.
2. Each equation describes the dependence of one variable in terms of the others, and contains a **stochastic error term** accounting for the influence of unobserved factors.
3. Independence assumptions on pairs of error terms are also specified in the model.

Structural Equation Models

1. The relationships among a set of observed variables are expressed by **linear equations**.
2. Each equation describes the dependence of one variable in terms of the others, and contains a **stochastic error term** accounting for the influence of unobserved factors.
3. Independence assumptions on pairs of error terms are also specified in the model.

Structural Equation Models

1. The relationships among a set of observed variables are expressed by **linear equations**.
2. Each equation describes the dependence of one variable in terms of the others, and contains a **stochastic error term** accounting for the influence of unobserved factors.
3. Independence assumptions on pairs of error terms are also specified in the model.

GAUSSIAN STRUCTURAL EQUATION MODELS

Let $G = (V, D, B)$ be a graph with vertex set $V = \{1, 2, \ldots, m\}$, a set of **directed edges** D, and a set of **bidirected edges** B. Assume the subgraph of directed edges is acyclic and topologically ordered.

Let PD_n denote the set of $m \times m$ **symmetric positive definite matrices**.

Let $PD(B) := \{\Omega \in PD_m : \omega_{ij} = 0 \text{ if } i \neq j \text{ and } i \leftrightarrow j \notin B\}$.

Let $\epsilon \sim \mathcal{N}(0, \Omega)$ such that $\Omega \in PD(B)$.

For each $i \to j \in D$ let $\lambda_{ij} \in \mathbb{R}$ be a **parameter**. For each $j \in V$ define

$$X_j = \sum_{i : i \to j \in D} \lambda_{ij} X_i + \epsilon_j.$$

GAUSSIAN STRUCTURAL EQUATION MODELS

$$X_j = \sum_{i:i \to j \in D} \lambda_{ij} X_i + \epsilon_j, \text{ for } j \in V$$

Let Λ be the strictly upper triangular matrix with $\Lambda_{ij} = \lambda_{ij}$ if $i \to j \in D$ and $\Lambda_{ij} = 0$ otherwise.

Since D is an acyclic digraph, the random vector $X = (X_1, \ldots, X_m)$ is well-defined centered multivariate normal distribution with **covariance matrix**

$$\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$$

Identification Problem

Decide whether the parameters in a structural model can be **determined uniquely** from the covariance matrix of the observed variables.

Equivalently, decide whether the following map is **injective**

$$\phi_G : (\Lambda, \Omega) \longrightarrow (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$$

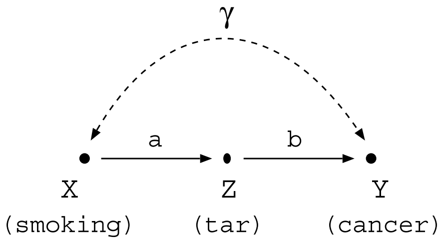
The identification of a model is important because, in general, no reliable quantitative conclusion can be derived from a non-identified model.

Example (Pearl 2000)

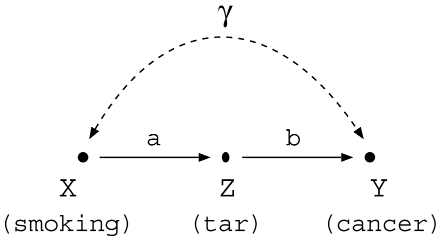
This model investigates the relations between **smoking** X and **lung cancer** Y, taking into consideration the **amount of tar** Z deposited in a person's lungs, and allowing for unobserved factors to affect both smoking X and cancer Y.

$$
\begin{aligned}
X &= \varepsilon_1 \\
Z &= aX + \varepsilon_2 \\
Y &= bZ + \varepsilon_3 \\
\mathrm{cov}(\varepsilon_1, \varepsilon_2) &= 0 \\
\mathrm{cov}(\varepsilon_2, \varepsilon_3) &= 0 \\
\mathrm{cov}(\varepsilon_1, \varepsilon_3) &= \gamma
\end{aligned}
$$

where $\varepsilon_i \sim \mathcal{N}(0, \omega_i)$.



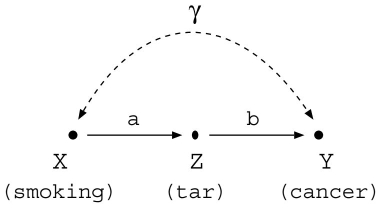
EXAMPLE (PEARL 2000)



$$\Sigma = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}$$

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \omega_1 & a\omega_1 & ab\omega_1 + \gamma \\ a\omega_1 & a^2\omega_1 + \omega_2 & a^2b\omega_1 + b\omega_2 + a\gamma \\ ab\omega_1 + \gamma & a^2b\omega_1 + b\omega_2 + a\gamma & a^2b^2\omega_1 + b^2\omega_2 + \omega_3 + 2ab\gamma \end{bmatrix}$$

EXAMPLE (PEARL 2000)



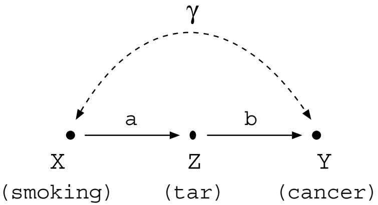
$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

$$a = \frac{\sigma_{12}}{\sigma_{11}} \qquad b = \frac{\sigma_{12}\sigma_{13} - \sigma_{11}\sigma_{23}}{\sigma_{12}^2 - \sigma_{11}\sigma_{22}}$$

$$\omega_1 = \sigma_{11} \qquad \omega_2 = \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{11}} \qquad \gamma = \frac{\sigma_{11}\sigma_{12}\sigma_{23} - \sigma_{11}\sigma_{13}\sigma_{22}}{\sigma_{12}^2 - \sigma_{11}\sigma_{22}}$$

$$\omega_3 = \frac{1}{(\sigma_{12}^2 - \sigma_{11}\sigma_{22})^2}(\sigma_{12}^2\sigma_{13}^2\sigma_{22} - 2\sigma_{12}^3\sigma_{13}\sigma_{23} + 2\sigma_{11}\sigma_{12}^2\sigma_{23}^2 - \sigma_{11}^2\sigma_{22}\sigma_{23}^2 + \sigma_{12}^4\sigma_{33} - 2\sigma_{11}\sigma_{12}^2\sigma_{22}\sigma_{33} + \sigma_{11}^2\sigma_{22}^2\sigma_{33})$$

EXAMPLE (PEARL 2000)



$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \omega_1 & a\omega_1 & ab\omega_1 + \gamma \\ a\omega_1 & a^2\omega_1 + \omega_2 & a^2 b\omega_1 + b\omega_2 + a\gamma \\ ab\omega_1 + \gamma & a^2 b\omega_1 + b\omega_2 + a\gamma & a^2 b^2\omega_1 + b^2\omega_2 + \omega_3 + 2ab\gamma \end{bmatrix}$$

ϕ_G IS GLOBALLY IDENTIFIABLE

The map $\phi_G : (\Lambda, \Omega) \longrightarrow (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}$ is injective **on the entire possible domain of definition.**

Global Identifiability

A directed graph D with at least two nodes is an **arborescence** converging to node i if i is the unique **sink** of D.

Given a mixed graph $G = (V, D, B)$ and a subset $A \subset V$, the **mixed subgraph induced by** A, denoted by $G_A = (A, D_A, B_A)$ is the graph containing all directed and bidirected edges whose endpoints are in A.

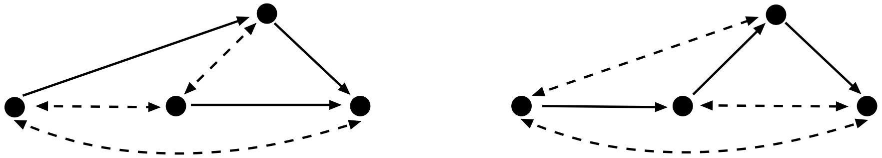
Theorem (Drton, Foygel, Sullivant)

The parametrization ϕ_G fails to be injective if and only if there is an induced subgraph G_A whose directed part (A, D_A) contains a converging arborescence and whose bidirected part (A, B_A) is connected.

The mixed graph $G = (V, D, B)$ is **simple**, or **bow-free** if $D \cap B = \emptyset$.

COROLLARY

*Suppose the map ϕ_G given by an acyclic mixed graph G is **injective**. Then G is a **simple**.*



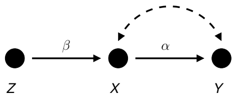
The two unlabeled simple graphs on four nodes with non-injective parametrization.

Generically Identifiable

Theorem (Brito, Perl)

Any simple graph G is generically identifiable.

$$\alpha = \frac{\sigma_{13}}{\sigma_{12}}$$
$$\beta = \frac{\sigma_{12}}{\sigma_{11}}$$
$$\omega_{XY} = \frac{\sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{22}}{\sigma_{12}}$$

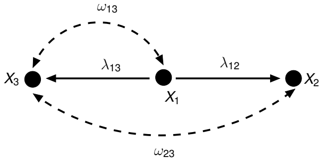


The parameter α is identified as long as $\sigma_{12} \neq 0$. The **instrumental variable** graph is not globally identified but **generically identified**.

Ω PARAMETERS

$$\Omega = (I - \Lambda)^T \Sigma (I - \Lambda).$$

So if the parameters in Λ are generically identifiable then the parameters in Ω are generically identifiable.



The parameter ω_{23} is identified, but λ_{13} is not identified.

$$\omega_{23} = \frac{\sigma_{11}\sigma_{23} - \sigma_{12}\sigma_{13}}{\sigma_{11}}$$

Approaches to the Identification Problem

Algebraic manipulation of the equations defining the model.

1. The method of path coefficients (Wright, 1934)
2. The rank and order criteria (Fisher, 1966)
3. Block recursive models (Fisher, 1966; Rigdon 1995)

Graphical Methods.

1. Single door criterion (Pearl, 2000)
2. Instrumental variables (Bowden and Turkington, 1984)
3. Back door criterion for total effects (Pearl, 2000)
4. G-criterion (Brito, 2006)
5. Graphical methods introduced by Tian (2004; 2005; 2007; 2009)
6. Recanting witness criterion for path-specific effects (Avin, Shpitser and Pearl, 2005)

Algebro-Geometric Approach

It remains unclear if these criteria (or combinations of the criteria) are **necessary and sufficient** to decide whether or not parameters are generically identifiable in a general mixed graph.

Introduce a **general** algebraic framework for performing identifiability computations for graphical models.

- Capable of testing direct effects, total effects, path-specific effects, error variances and covariances (Ω parameters).
- Provides certificates if a given parameter is not identifiable.
- Capable of detecting **algebraic** *d*-identifiable parameters (*d*-to-one parameters).

Let k be a field. An **affine variety** is the common zero locus of polynomials $f_1, \ldots, f_r \in k[x_1, \ldots, x_n]$.

$$V(f_1, \ldots, f_r) = \{f_1 = 0, f_2 = 0, \ldots, f_r = 0\}$$

$$V(1) = \emptyset$$
$$V(0) = k^n$$

For **linear polynomials** f_i, $V(f_1, \ldots, f_r)$ is the solution space of an inhomogeneous system of linear equations. This variety is described parametrically applying **Gauss Algorithm**.

Ideals

Given $f_1, \ldots, f_r$ in $k[x_1, \ldots, x_n]$,

$$\langle f_1, \ldots, f_r \rangle = \{\sum_{i=i}^{n} h_i f_i \mid h_1, \ldots, h_n \in k[x_1, \ldots, x_n]\}$$

is the **ideal** generated by $f_1, \ldots, f_r$.

Theorem (Hilbert Basis Theorem)

If $I \subset k[x_1, \ldots, x_n]$ is an ideal, there exists $f_1, \ldots, f_r$ such that

$$\langle f_1, \ldots, f_r \rangle = I.$$

Rational Parametrization

Let $p_1(t_1,\dots,t_d),\dots,p_n(t_1,\dots,t_d) \in k[t_1,\dots,t_d]$. The set

$$S = \{(p_1(a_1,\dots,a_d),\dots,p_n(a_1,\dots,a_d)) \in k^n \mid (a_1,\dots,a_d) \in k^d\}$$

is called a **rational parametrization (r.p.)**.

$$I(S) = \{g \in k[x_1,\dots,x_n] \mid g(a_1,\dots,a_n) = 0 \text{ for all } (a_1,\dots,a_n) \in S\}$$

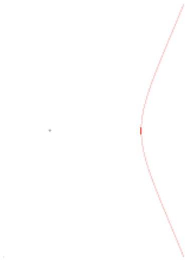
$I(S)$ is the **ideal of polynomial functions vanishing on S**.

Theorem

If S is a r.p. and $I(S) = \langle f_1,\dots,f_r\rangle$ then S and $V(f_1,\dots,f_r)$ differ by a set of dimension less than the dimension of S.

CUBIC PLANE CURVE

Let $S = \{(t^2 + 1, t^3 + t) \mid t \in \mathbb{R}\}$. Then $I(S) = \langle y^2 - x^3 - x^2 \rangle$



For any ideal $I \subset k[x_1, \ldots, x_n]$ we consider the **elimination ideal**

$$I_m = I \cap k[x_{m+1}, \ldots, x_n]$$

and the **projection** $\pi_m : k^n \longrightarrow k^{n-m}$ given by

$$\pi_m(a_1, \ldots, a_n) = (a_{m+1}, \ldots, a_n)$$

THEOREM

If k is ***algebraically closed****, then $\overline{\pi_m(V(I))} = V(I_m)$.*

THEOREM

If $G = \{g_1, \ldots, g_r\}$ is a ***Gröbner basis*** *of $I \subset k[x_1, \ldots, x_n]$ with respect to the lexicographic ordering, then*

$$G_m = G \cap k[x_{m+1}, \ldots, x_n]$$

is a Gröbner basis of $I_m \subset k[x_{m+1}, \ldots, x_n]$ w.r.t. lex.

THEOREM

Suppose k is an ***infinite field*** *and we are given a rational map* $\phi : k^m \setminus Z \longrightarrow k^n$ *given by*

$$(t_1, \ldots, t_m) \longmapsto \Big(\frac{f_1(t)}{g_1(t)}, \ldots, \frac{f_n(t)}{g_n(t)}\Big)$$

with f_i *and* $g_i \in k[t_1, \ldots, t_m]$ *and* $Z = V(g_1 \cdot g_2 \cdots g_n)$. *Let*

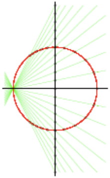
$$I = \langle g_1 x_1 - f_1, \ldots, g_n x_n - f_n, 1 - gs \rangle \subset k[s, t_1, \ldots, t_m, x_1, \ldots, x_n]$$

Then

$$\overline{\mathrm{Im}(\phi)} = V(I_{m+1})$$

Implicitization of the unit Circle

Consider the parametrization of the unit circle given by the 2nd intersection point of $y = t(x+1)$ with the circle:



$$t \longmapsto \left(\frac{1-t^2}{t^2+1}, \frac{2t}{t^2+1}\right)$$

Let $I = \left\langle (t^2+1)x - (1-t^2), (t^2+1)y - 2t, 1 - (t^2+1)^2 s \right\rangle$. The set

$$\{x^2+y^2-1, ty+x-1, tx+t-y, s-\frac{1}{2}x-\frac{1}{2}\}$$

is a Gröbner basis with respect to lex $s > t > x > y$. Hence

$$I_2 = \left\langle x^2+y^2-1 \right\rangle$$

.

Identifiable Parameters

Let $\Theta \subseteq \mathbb{R}^d$ be a full dimensional **parameter set**.

Let $\phi_1, \ldots, \phi_n \in \mathbb{R}[t_1, \ldots, t_d]$, and $\Phi : \Theta \to \mathbb{R}^n$ be the function defined by $\Phi(\theta) = (\phi_1(\theta), \ldots, \phi_n(\theta))^T$.

A **parameter** is a polynomial function $u : \Theta \to \mathbb{R}$ which is not constant on Θ.

The parameter u is **identifiable** if there exists a map $\Psi : \mathbb{R}^n \to \mathbb{R}$ such that $u(\theta) = \Psi \circ \Phi(\theta)$ for all $\theta \in \Theta$.

The parameter u is **generically identifiable** if there exists a map $\Psi : \mathbb{R}^n \to \mathbb{R}$ and a dense open subset $\mathcal{O}$ of Θ such that $u(\theta) = \Psi \circ \Phi(\theta)$ for all $\theta \in \mathcal{O}$.

Algebraic Approach

Given $\Phi : \Theta \to \mathbb{R}^n$ defined by $\Phi(\theta) = (\phi_1(\theta), \ldots, \phi_n(\theta))^T$, we want to check if the **parameter** u is (generically) **identifiable**.

The **vanishing ideal** of $S \subseteq \mathbb{R}^n$ is the set

$$\mathcal{I}(S) := \{g \in \mathbb{R}[p_1, \ldots, p_n] : g(\mathbf{a}) = 0 \text{ for all } \mathbf{a} \in S\}.$$

Let $\tilde{\Phi} = (u, \phi_1, \ldots, \phi_n)^T : \Theta \to \mathbb{R}^{d+1}$.

Let $\mathbb{R}[q, \mathbf{p}]$ be the polynomial ring with **one extra indeterminate** corresponding to the parameter function u.

Let $\mathcal{I}(\tilde{\Phi}(\Theta))$ be the vanishing ideal of the image.

Parameter Identifiability

Theorem

Suppose that $g(q,\mathbf{p}) \in \mathcal{I}(\tilde{\Phi}(\Theta))$ is a polynomial such that q appears in this polynomial, $g(q,\mathbf{p}) = \sum_{i=0}^{d} g_i(\mathbf{p})q^i$ and $g_d(\mathbf{p})$ does not belong to $\mathcal{I}(\Phi(\Theta))$.

1. *If g is* ***linear*** *in q, $g = g_1(\mathbf{p})q - g_0(\mathbf{p})$ then u is* ***generically identifiable*** *by the formula $u = \frac{g_0(\mathbf{p})}{g_1(\mathbf{p})}$. If, in addition, $g_1(\mathbf{p}) \neq 0$ for $\mathbf{p} \in \Phi(\Theta)$ then u is* ***identifiable****.*
2. *If g has* ***higher degree*** *d in q, then u is* ***algebraically d-identifiable*** *(may or might not be identifiable).*
3. *If no such polynomial g exists then the parameter u is* ***not generically identifiable****.*

Theorem

Of the 64 mixed graphs on three vertices,

- *there are exactly 31 graphs that are* ***generically identifiable*** *and 33 graphs that are* ***not generically identifiable****.*
- *The single-door criterion and instrumental variables form a* ***complete method*** *to generically identify direct causal effects for SEM models on three variables.*

Computational Results

Theorem

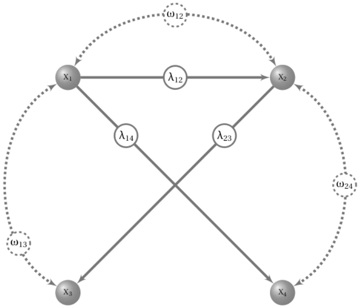
Of the 4096 mixed graphs on four variables

- *exactly 1246 are **generically identifiable**, 6 are **algebraically 2-identified**, and 2844 are **not generically identifiable**.*
- *Of the 1246 generically identifiable models, exactly 1093 are generically identified by the single-door and instrumental variables criteria and the remaining 153 generically identified models contain direct causal effect parameters **only identified by the algebraic method**.*
- *There are exactly 729 bow-free models, each generically identified by the single-door criterion.*

Structural Equation Models Web Site

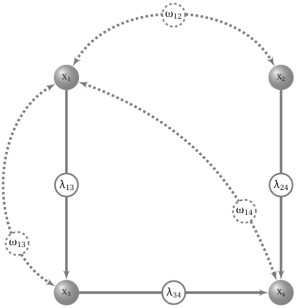
http://www.shsu.edu/research/graphicalmodels/

- Show that λ_{12} is generically identifiable.
- Do you know of a graphical criterion that can identify this parameter?
- Show that G is generically identifiable.



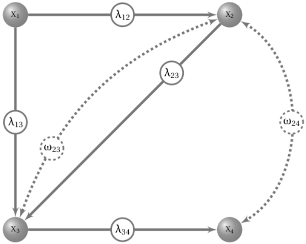
Challenge 2

- Show that λ_{24} and λ_{34} are generically identifiable.
- Do you know of a graphical criterion that can identify these parameters?
- Show that G is generically identifiable.



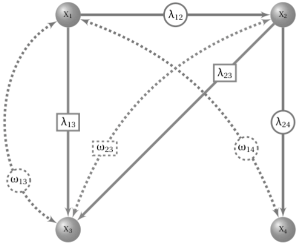
CHALLENGE 3

- Show that λ_{13} and λ_{23} are generically identifiable.
- Do you know of a graphical criterion that can identify these parameters?
- Show that G is generically identifiable.



Challenge 4

- Show that λ_{13} and λ_{23} are **not** generically identifiable.
- Show that the **total effect** of X_1 on X_3: $\lambda_{12}\lambda_{23} + \lambda_{13}$ is generically identifiable.
- This total effect is **not** identified by the **Back-Door criterion**. Is there a graphical criterion that identifies this parameter?



Challenge 5

Table: Algebraically 2-identified SEMs on four variables.

Directed edges	Bidirected edges
$1 \to 2, 2 \to 3, 3 \to 4$	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 1 \leftrightarrow 4$
$1 \to 2, 2 \to 3, 2 \to 4$	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 1 \leftrightarrow 4$
$1 \to 2, 1 \to 4, 2 \to 3$	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 1 \leftrightarrow 4$
$1 \to 2, 1 \to 3, 3 \to 4$	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 1 \leftrightarrow 4$
$1 \to 2, 1 \to 3, 2 \to 4$	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 1 \leftrightarrow 4$
$1 \to 2, 1 \to 3, 1 \to 4$	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 1 \leftrightarrow 4$

Is there a combinatorial description of these mixed graphs?