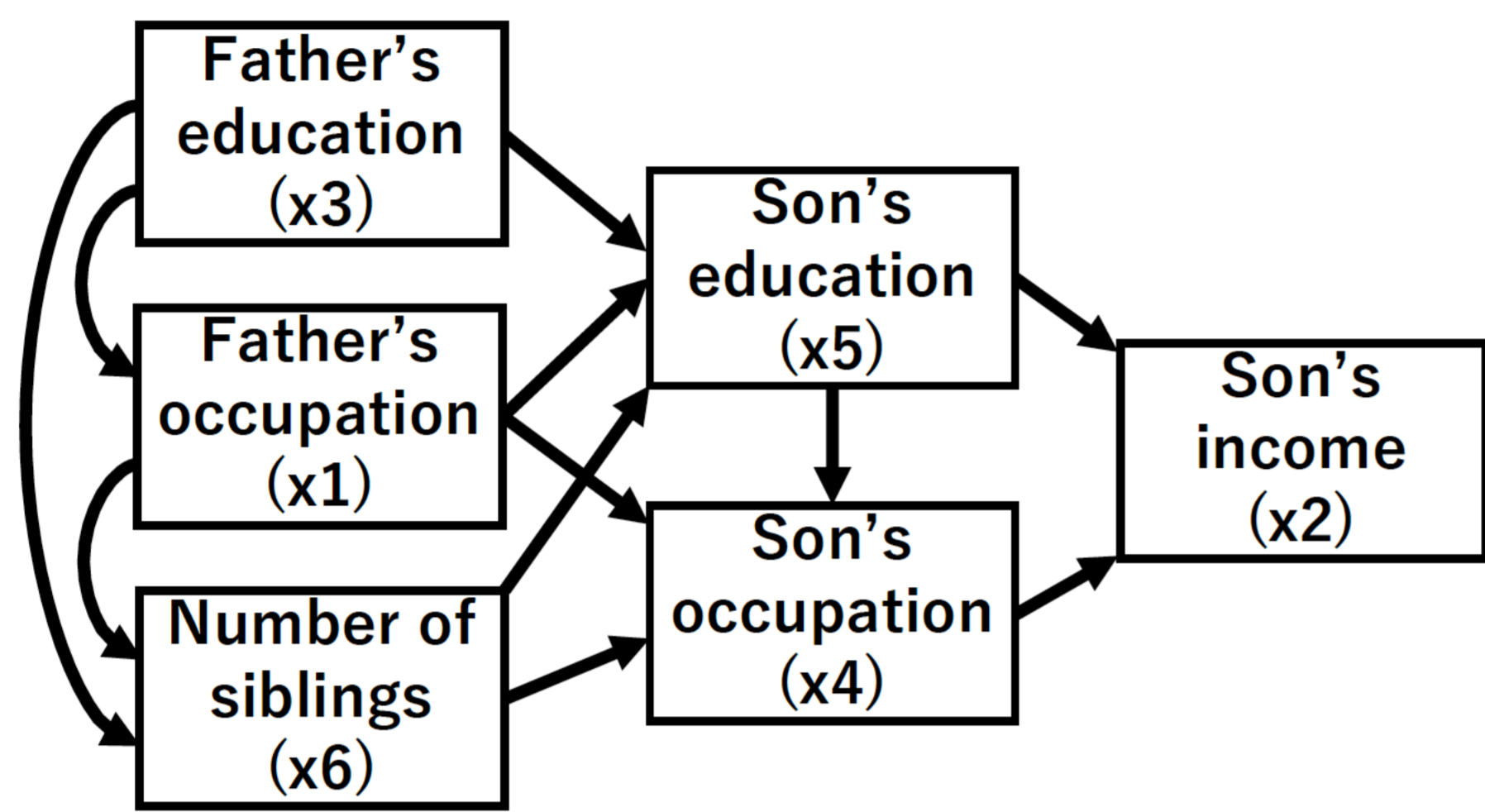


未観測共通原因を含むデータからの因果グラフ推定 (特別研究員 前田 高志ニコラス)

観察データからの因果グラフの推定

因果推定には、対照実験を行うことが最も効果的だが、実験が倫理面・コスト面で難しい場合もある
→観察データから因果関係を推定する



因果グラフの例

因果関数モデルによる因果推定

Additive Noise Models: $X_i = f(\mathbf{Pa}X_i) + N_i$
(X_i : 観測変数、 $\mathbf{Pa}X_i$: X_i の原因変数、 N_i : 外生変数)

因果関数モデルによる因果推定
 $Y=f(X)$ が X と独立となる関数 f が存在し、 $X=g(Y)$ が Y と独立となる関数 g が存在しないとき、 X は Y の原因

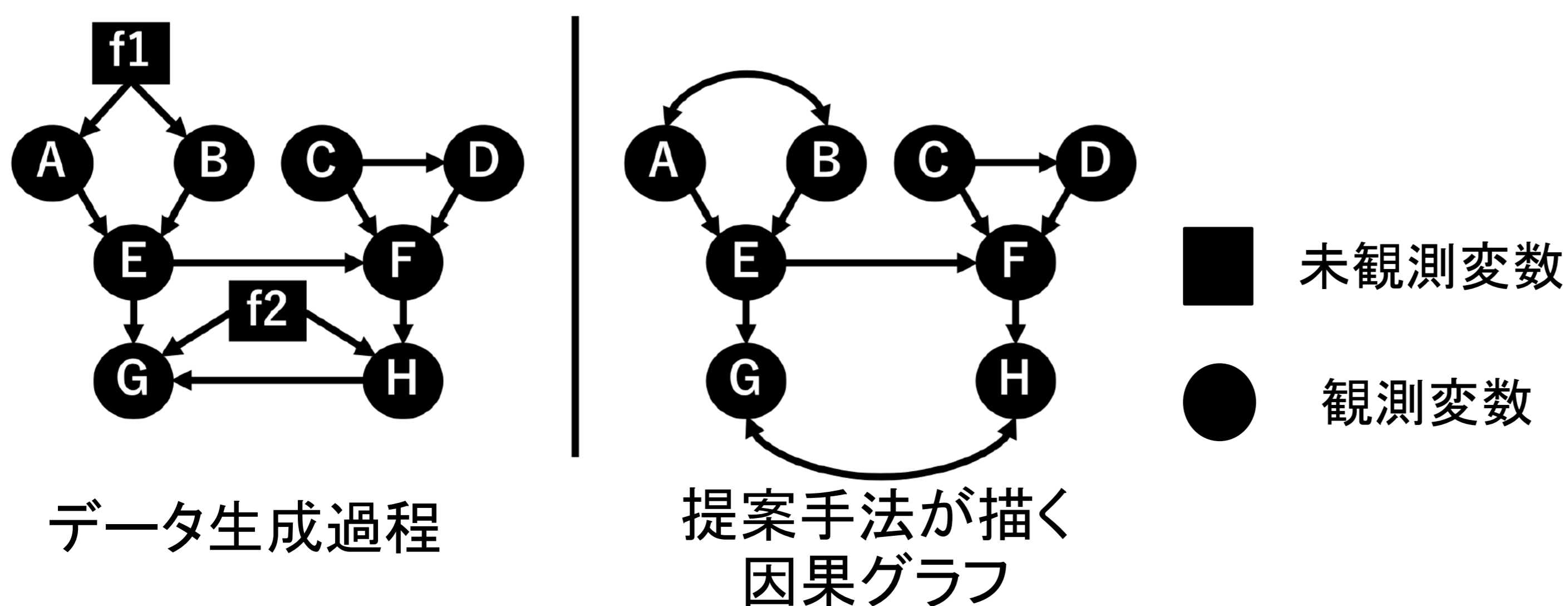
未観測共通原因

2つ以上の観測変数に影響を与える未観測共通原因が存在すると、その観測変数間の因果関係が分からず、全体の因果グラフも推定困難になってしまう

1. 未観測共通原因を含むデータからの因果グラフ推定

T. N. Maeda, S. Shimizu (AISTATS2020)

- 各観測変数の原因を因果関数モデルにより推定
- 各変数の各原因が直接的原因なのか間接的原因なのかを判定
- 互いに因果関係がないが、非独立な変数ペアに未観測共通原因があると推定



社会科学のための因果探索法の開発 (特別研究員 李超)

因果探索とは

ランダム化実験が難しいとき、調査観察データによる因果仮説の探索手法である。最近、非ガウス性を利用するLINGAMモデルが注目されている。

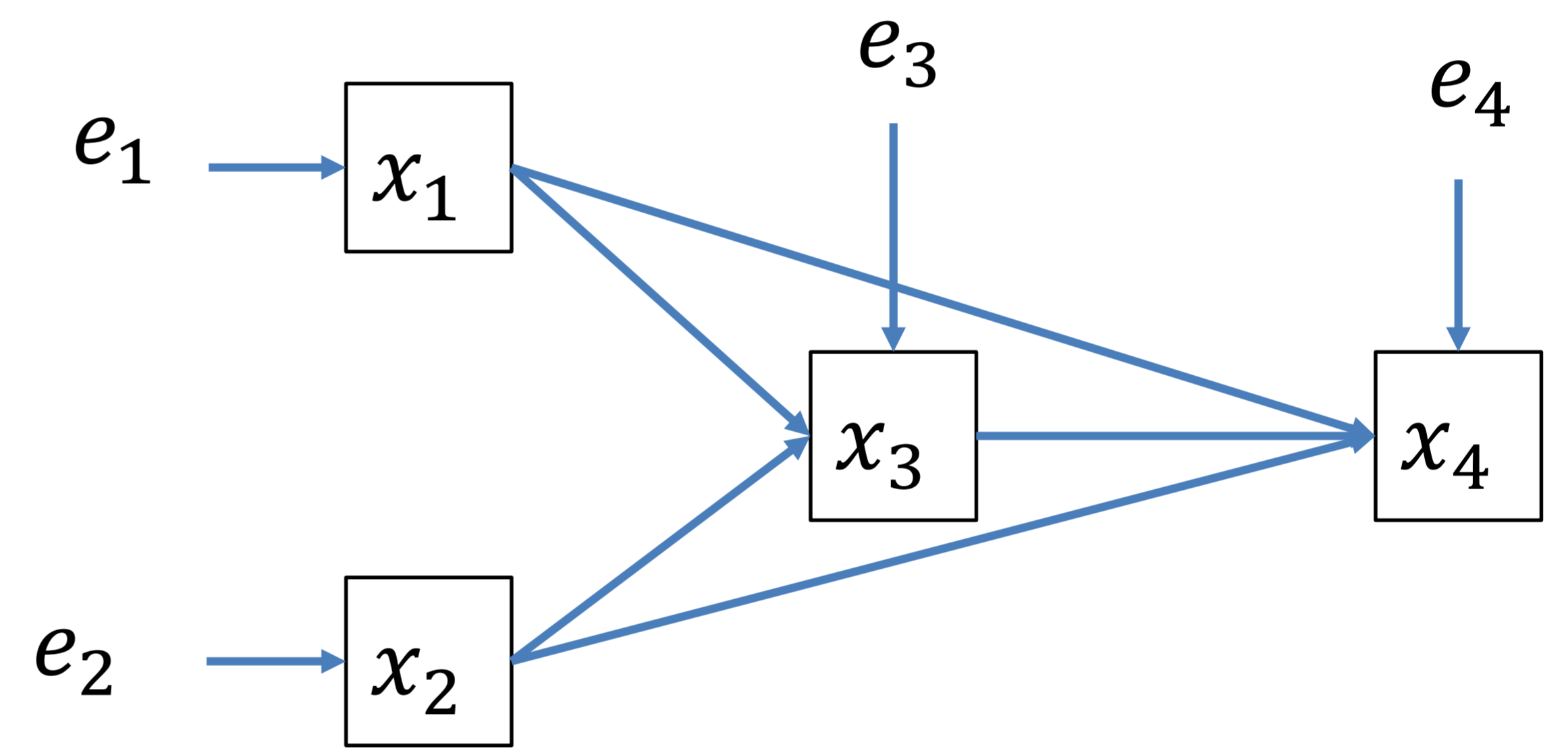
LINGAMモデル

Linear Non-Gaussian Acyclic Model (LiNGAM) (shimizu et al., 2006)

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad (x = Bx + e)$$

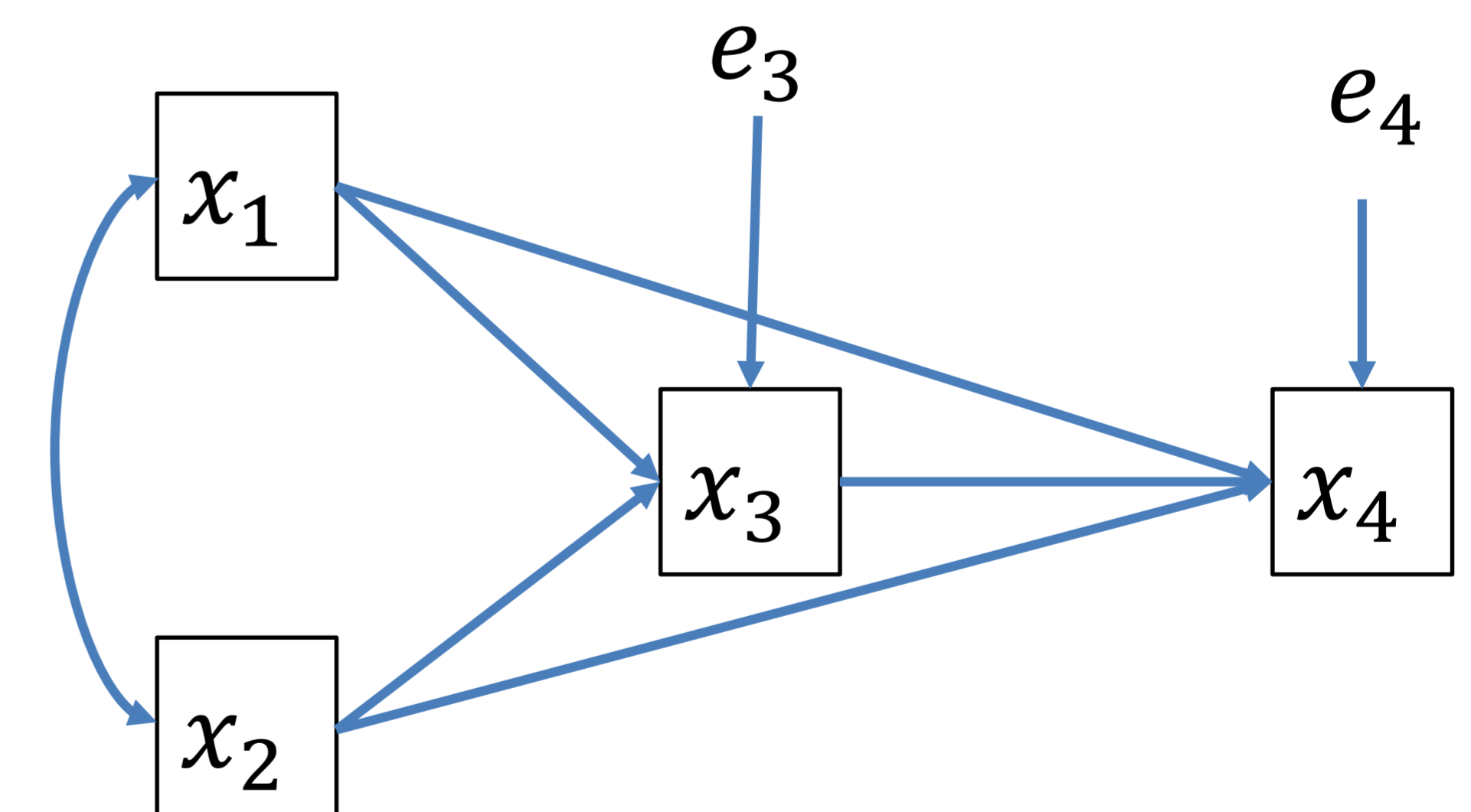
行列表現

- 非巡回
- 非ガウス誤差 e_i
- e_i は互いに**独立** (潜在共通原因なし)



LINGAMモデルは**外生変数** x_1, x_2 が互いに**独立**と仮定する。しかし、社会科学ではよく用いられる因果解析では外生変数同士には相関をつける。そこで、以下のように拡張したLINGAMモデルとその推定法を提案する

拡張したLINGAMモデル



- 内生変数部分の因果グラフは非循環
- $x_i = \sum_{j \neq i} b_{ij} x_j + e_i$ (x_i は内生変数)
- 非ガウス誤差 e_i
- 外生変数(x_1, x_2)は多変量T分布に従う

モデルの高速探索手法

逐次的にsink変数(どの変数の原因にもならない)である可能性がもっとも大きい変数(BICスコアで評価)を決める手法