

経時データにおける非ガウス性を用いた因果構造探索

Learning causal structure of longitudinal data with a use of non-Gaussianity

門脇 健人 清水 昌平 鷲尾 隆
Kento Kadowaki Shohei Shimizu Takashi Washio

大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

Recently, there is a growing need for statistical learning of causal structures in data with many variables. A structural equation model called Linear Non-Gaussian Acyclic Model (LiNGAM) has been extensively studied to estimate causal structures in data. In LiNGAM, we can uniquely estimate the causal structure. The key assumptions are that external influences are independent and follow non-Gaussian distributions. In this paper, we consider learning causal structures in longitudinal data that collects samples over a time period. In previous studies on LiNGAM, there was no model specialized to model longitudinal data. Therefore, we propose a new model called longitudinal LiNGAM and new estimation method using the longitudinal structure and non-Gaussianity of data. The new approach requires weaker conditions than previous methods.

1. はじめに

データマイニングの分野では、観測したデータに潜む因果構造を推定するための統計的手法について近年盛んに研究がなされている。因果構造の推定とは、具体的には観測されたある現象が他の現象にいかに関与を及ぼしているかを知ることである。このような分野は因果推論分野と呼ばれ、推定された因果構造をグラフィカルに表現できるのが特徴であり、それ故に様々な分野で応用研究がなされている。

本研究では時間の経過とともに繰り返し観測された経時データと呼ばれる観測データの因果構造を正しく推定することを主なテーマとする。従来は経時データの因果構造を推定する際、既存の非ガウス線形構造方程式モデルである LiNGAM 分析をアドホックに適用した手法 [Shimizu 12] や、構造型自己回帰モデルのフレームワークで LiNGAM モデルと自己回帰モデルを組み合わせた手法により推定を行っていた [Swanson 97, Hyvärinen 10]。しかし、従来の手法においては、経時データの因果構造が時間の経過とともに変化する場合や過去の変数からの影響が存在する場合に正しく推定出来ないという問題があった。本研究ではそのような問題点を改善し、経時データの特徴を正しく記述するモデルを提案することで、経時データの因果構造に対する推定精度の向上を試みる。これに対し人工データを用いて性能評価を行い、その結果について考察する。

2. 経時データ

経時データとは、複数のサンプルにおいて時間の推移とともに各時点で観測した値の集合であり、その時点の順序に意味を持つデータの集合である。ここで時点の数は有限かつ離散値であり、時点間の間隔は観測者の任意の間隔で与えられているものとする。

時間の経過とともに観測された各変数を $x_i(t)$ ($i = 1, \dots, p, t = 1, \dots, T$) と定義する。ある時点 t である一つのサンプル m において観測された各変数 $x_i(t)$ を p 個並べた列ベクトルを $\mathbf{x}_m(t)$ とすると、 $\mathbf{x}_m(t)$ に関する一つの時系列デー

タは $\mathbf{x}_m(1), \dots, \mathbf{x}_m(T)$ ($m = 1, \dots, n$) で表される。また、ある時点 t におけるサンプル $m = 1, \dots, n$ の列ベクトル $\mathbf{x}_m(t)$ を束ねた $p \times n$ 行列を $\mathbf{X}(t) = [\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)]$ とする。経時データと呼ばれる観測データ群は T 個の行列 $\mathbf{X}(t)$ をその時間順序に束ねた $p \times n \times T$ のデータテンソル \mathbf{X} で表される。ここで、 m 番目のサンプルを一つ指定すると一つの時系列データ $\mathbf{x}_m(1), \dots, \mathbf{x}_m(T)$ が得られ、時点 t を一つ指定するとその時点で観測されたデータ行列 $\mathbf{X}(t)$ が得られる。

3. LiNGAM モデル

LiNGAM (Linear Non-Gaussian Acyclic Model) モデル [Shimizu 06, Shimizu 11] は、データが非ガウス分布に従うことを仮定した上で、構造方程式モデルに線形非巡回性を持たせたモデルである。観測変数 x_i ($i = 1, \dots, n$) が LiNGAM モデルに従って生成されるとき、 x_i は次式で表される：

$$x_i = \sum_{k(j) < k(i)} b_{ij} x_j + e_i \quad (1)$$

ただし、 $k(i)$ は変数の因果的順序、 e_i は非ガウス分布に従う外乱、 b_{ij} は定数の係数である。式 (1) が表す x_i の生成過程は有向非巡回グラフ (Directed Acyclic Graph, DAG) で表現できる。このとき $B = [b_{ij}]$ は有向非巡回グラフの隣接行列にあたり、 b_{ij} は変数 x_j から x_i への影響の強さを表している。式 (1) の行列表記は次式で与えられる：

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e} \quad (2)$$

今、 \mathbf{x} の生成過程が非巡回性であることを仮定しているため、変数の因果的順序 $k(i)$ に従って変数 \mathbf{x} の順序を並び替えると \mathbf{B} は狭義下三角行列^{*1} となる [Bollen 89]。LiNGAM 分析によって、データ行列 \mathbf{X} から因果的順序 $k(i)$ と隣接行列 \mathbf{B} を推定することができる [Shimizu 06]。LiNGAM 分析の例として、近年 DirectLiNGAM 分析 [Shimizu 11] が提案されている。本研究では DirectLiNGAM 分析を用いて因果構造を推定する。

連絡先: 門脇 健人, 大阪大学 産業科学研究所, 567-0047 大阪府 茨木市美穂ヶ丘 8-1, kadowaki@ar.sanken.osaka-u.ac.jp

*1 対角成分が 0 であるような下三角行列。

4. 従来手法

4.1 SVAR モデル

経時データの因果構造を求めるために、一つの $p \times T$ 時系列データを入力とした以下の構造型自己回帰モデル (以下 SVAR モデル)[Swanson 97] を用いることが多い。まず SVAR モデル [Swanson 97] について述べる前に、自己回帰モデルと呼ばれる古典的な時系列モデル [?] について説明する。時間の経過とともに観測された各変数を $x_i(t), i = 1, \dots, p, t = 1, \dots, T$ とし、各変数 $x_i(t)$ の平均は 0 としても一般性は失われない。時点 t における変数を並べた列ベクトルを $\mathbf{x}(t)$ とすると、自己回帰モデルは以下で定義される。

$$\mathbf{x}(t) = \sum_{\tau=1}^l \mathbf{B}_\tau \mathbf{x}(t-\tau) + \mathbf{e}(t) \quad (3)$$

ここで、パラメータ l は何時点前までの影響を考慮するかを表し、 $\mathbf{B}_\tau (\tau = 1, \dots, l)$ は $p \times p$ 行列である。 p 次元ベクトル $\mathbf{e}(t)$ は外乱を表す。

実際の観測データでは、変数 $x_i(t)$ 間の影響は同時点間及び過去の時点からの影響を同時に含むことが多い [Swanson 97]。そこで、過去の時点からの影響を自己回帰モデル (3) で表し、同時点間の影響を構造方程式モデルで表した SVAR モデル [Swanson 97] が提案された。

SVAR モデル：

$$\mathbf{x}(t) = \sum_{\tau=0}^l \mathbf{B}_\tau \mathbf{x}(t-\tau) + \mathbf{e}(t) \quad (4)$$

\mathbf{B}_τ は $p \times p$ 行列であり、 $\tau (\tau > 0)$ 時点前からの変数 x_i からの影響の強さを表し、 $\tau = 0$ の時は同時点の影響を表す行列である。[Hyvärinen 10] では、SVAR モデルの因果構造を従来の LiNGAM 解析手法を用いて推定するために外乱 $\mathbf{e}(t)$ に対して以下の仮定を加える。

1. $e_i(t)$ は全ての時点、全ての変数と互いに独立。
2. $e_i(t)$ は非ガウス分布に従う。
3. 同時点間の影響を表す行列 \mathbf{B}_0 は非巡回グラフに従う。

以上の仮定は LiNGAM モデルの仮定 [Shimizu 06] と同様のものである。新たに LiNGAM の仮定を加えた SVAR モデルを AR-LiNGAM モデル [Hyvärinen 10] として次節でその推定手法を述べる。

4.2 AR-LiNGAM モデルの推定

モデル (4) の外乱 $e_i(t)$ は互いに独立かつ非ガウス分布であるという仮定、そして残差と元の観測変数の因果的順序は等しいという関係により過去の影響を取り除いた残差に LiNGAM 分析を適用することができる。これにより、モデル (4) における同時変数間の影響 \mathbf{B}_τ を求めることが可能となる。

AR-LiNGAM モデルの推定における基本的な考えは、最小二乗回帰などの古典的な手法を用いて自己回帰行列 \mathbf{M}_τ を推定し、そこから残差 $\mathbf{n}(t)$ に DirectLiNGAM などの LiNGAM 分析を適用して同時点間の影響 \mathbf{B}_0 を得る、というものである。これにより、過去の変数からの影響と同時点間の変数の影響を同時に求めることができる。以下に、アルゴリズム 1 として AR-LiNGAM モデルの具体的な推定手法を示す [Hyvärinen 10]。

アルゴリズム 1: AR-LiNGAM アルゴリズム

入力: p 変数, T 時点の $p \times T$ データ行列 \mathbf{X}_{AR} , ($p \ll T$)

1. 入力データに対し以下の自己回帰モデルを適用する。

$$\mathbf{x}(t) = \sum_{\tau=1}^l \mathbf{M}_\tau \mathbf{x}(t-\tau) + \mathbf{n}(t)$$

ここで、 $\mathbf{n}(t)$ は AR モデルにおける誤差項である。AR モデルに対し、最小二乗法や最尤法など共分散に基づく既存の回帰手法を適用し、 $\mathbf{M}_\tau, \tau > 0$ を得る。ここで、最小二乗法を用いて推定された自己回帰行列を $\hat{\mathbf{M}}_\tau$ とする。

2. 式 (1) より、以下の残差 $\mathbf{n}(t)$ を計算する。

$$\hat{\mathbf{n}}(t) = \mathbf{x}(t) - \sum_{\tau=1}^l \hat{\mathbf{M}}_\tau \mathbf{x}(t-\tau)$$

3. 残差 $\mathbf{n}(t)$ は LiNGAM モデルに従うので、DirectLiNGAM などの LiNGAM 解析手法を用いて同時点間の変数の影響を表す \mathbf{B}_0 を推定する。

$$\hat{\mathbf{n}}(t) = \mathbf{B}_0 \hat{\mathbf{n}}(t) + \mathbf{e}(t)$$

4. 最後に過去の変数からの影響 $\mathbf{B}_\tau, \tau > 0$ を以下により求める。

$$\hat{\mathbf{B}}_\tau = (\mathbf{I} - \hat{\mathbf{B}}_0) \hat{\mathbf{M}}_\tau \quad (\tau > 0)$$

出力として、同時点の変数間の影響 $\hat{\mathbf{B}}_0$ と過去の変数の影響 $\hat{\mathbf{B}}_\tau (\tau > 0)$ を得る。

既存手法の AR-LiNGAM アルゴリズムでは、入力データを一つの時系列 \mathbf{X}_{AR} としたが、これだけでは従来の LiNGAM モデルより複雑な構造を持ちうる経時データの因果構造を推定するのに情報が少ない。故に AR-LiNGAM モデルでは、同時点間の変数の影響 \mathbf{B}_0 及び過去の変数からの影響 $\mathbf{B}_\tau (\tau > 0)$ は時点 t によらず一定という制約を入れざるを得なかった。以上のことから、既存研究では経時データの因果構造が時間的に変化する場合に正しく推定することが出来ないという問題がある。また、LiNGAM 分析をアドホックに用いて経時データの因果構造を推定する手法も考えられたが、そのような手法においても仮定が崩れているため過去の変数からの影響がある場合や因果構造が時間的に変化する場合には正しい推定が行えないという問題があった。

5. 提案手法

5.1 経時 LiNGAM モデル

SVAR モデルでは入力データをサンプルサイズ $n = 1$ の経時データに限定していたが、実際のデータでは多くのサンプルサイズで観測した経時データを得られる場合がある。サンプルサイズの大きい経時データを得ることによって、同時点間の変数の影響及び過去の変数からの影響は時点によらず一定という AR-LiNGAM モデルの制約を無くした新たなモデルを提案する。

定義 1 (経時 LiNGAM モデル). 経時データの定義から、 $p \times n \times T$ テンソルの時点 t における $p \times n$ 行列を $\mathbf{X}(t)$ とした時、経時 LiNGAM モデルは以下の式で表される：

$$\mathbf{X}(t) = \sum_{\tau=0}^l \mathbf{B}(t, t-\tau)\mathbf{X}(t-\tau) + \mathbf{e}(t) \quad (5)$$

式 (5) で表されるモデルは以下の特徴を持つ．

- $\mathbf{e}(t)$ は外乱項であり，各 $e_i(t)$ は非ガウス分布に従う．また，同時点の $e_i(t)$ 及び他のどの時点の $e_i(\tau)$ とも独立である．
- $\mathbf{B}(t, t-\tau)$ は $p \times p$ 行列であり， $\tau(\tau > 0)$ 時点前からの変数 $x_i(t-\tau)$ からの影響の強さを表し， $\tau = 0$ である行列 $\mathbf{B}(t, t)$ は同時点の影響を表す行列である．また， $\mathbf{B}(t, t-\tau)$ は時間差 τ だけでなく時点 t にも依存するので，因果構造が時間的に変化する場合も記述できる．
- l は時間差を表し，何時点前までの変数の影響を考えるかを意味する．
- 最初の観測時点 $t = 1$ における因果構造は潜在交絡変数 [Hoyer 08] の影響を受ける．潜在交絡変数の影響を受けている部分の推定については本研究の範囲を超えるので， $t = 1$ における因果構造は推定しない．潜在交絡変数とその影響が存在する場合の推定については，例えば [Hoyer 08] を参照されたい．

5.2 経時 LiNGAM アルゴリズム

経時 LiNGAM モデルの因果構造を推定する手法について説明する．はじめに，大まかな推定の手順を示す．式 (5) より

$$(\mathbf{I} - \mathbf{B}(t, t))\mathbf{X}(t) = \sum_{\tau=1}^l \mathbf{B}(t, t-\tau)\mathbf{X}(t-\tau) + \mathbf{e}(t)$$

これをさらに変形して

$$\begin{aligned} \mathbf{X}(t) - \sum_{\tau=1}^l (\mathbf{I} - \mathbf{B}(t, t))^{-1} \mathbf{B}(t, t-\tau)\mathbf{X}(t-\tau) \\ = (\mathbf{I} - \mathbf{B}(t, t))^{-1} \mathbf{e}(t) \end{aligned}$$

が導ける．ここで $\mathbf{X}(t) - \sum_{\tau=1}^l (\mathbf{I} - \mathbf{B}(t, t))^{-1} \mathbf{B}(t, t-\tau)\mathbf{X}(t-\tau)$ を $\mathbf{N}(t)$ *2 とする． $\mathbf{B}(t, t)$ は定義より狭義下三角行列であるから， $\mathbf{I} - \mathbf{B}(t, t)$ は正則である．よって $\mathbf{I} - \mathbf{B}(t, t)$ には常に逆行列が存在し，以下のように式変形できる．

$$\mathbf{N}(t) = (\mathbf{I} - \mathbf{B}(t, t))^{-1} \mathbf{e}(t) \iff \mathbf{N}(t) = \mathbf{B}(t, t)\mathbf{N}(t) + \mathbf{e}(t)$$

従って， $\mathbf{N}(t)$ は LiNGAM モデルに従う．ここで，行列 $(\mathbf{I} - \mathbf{B}(t, t))^{-1} \mathbf{B}(t, t-\tau)$ を $\mathbf{M}(t, t-\tau)$ として表すと，式 $\mathbf{X}(t) = \sum_{\tau=1}^l \mathbf{M}(t, t-\tau)\mathbf{X}(t-\tau) + \mathbf{N}(t)$ に従って目的変数を \mathbf{X} ，説明変数を $\mathbf{X}(t-1), \dots, \mathbf{X}(t-l)$ として最小二乗回帰などの古典的な手法を用いて $\mathbf{M}(t, t-\tau)$ を推定し，そこから過去の変数からの影響を取り除いた $\mathbf{N}(t)$ に DirectLiNGAM などの LiNGAM モデルの解析手法を適用して同時点間の影響 $\mathbf{B}(t, t)$ を推定する．ここで，行列 $\mathbf{M}(t, t-\tau)$ は時間差 τ と時点 t の双方に依存する．具体的なアルゴリズムは以下である．

アルゴリズム 2: 経時 LiNGAM アルゴリズム

入力: p 変数, サンプルサイズ n , T 時点の $p \times n \times T$ データテンソル \mathbf{X} , ($p \ll n$)

1. 入力データに対し以下のモデルを適用する．

$$\mathbf{X}(t) = \sum_{\tau=1}^l \mathbf{M}(t, t-\tau)\mathbf{X}(t-\tau) + \mathbf{N}(t)$$

ここで， $\mathbf{N}(t)$ は (1) のモデルにおける誤差項である．(1) 式モデルに対し， $\mathbf{X}(t)$ を目的変数， $\mathbf{x}(t-1), \dots, \mathbf{x}(t-l)$ を説明変数として最小二乗法や最尤法など共分散に基づく既存の回帰手法を適用し， $\mathbf{M}(t, t-\tau)$, $\tau > 0$ を得る．ここで，最小二乗法を用いて推定された自己回帰行列を $\hat{\mathbf{M}}(t, t-\tau)$ とする．

2. 式 (1) より，以下の残差 $\mathbf{N}(t)$ を計算する．

$$\hat{\mathbf{N}}(t) = \mathbf{X}(t) - \sum_{\tau=1}^l \hat{\mathbf{M}}(t, t-\tau)\mathbf{X}(t-\tau)$$

3. 残差 $\mathbf{N}(t)$ は LiNGAM モデルに従うので，DirectLiNGAM などの LiNGAM 解析手法を用いて同時点間の変数の影響を表す $\mathbf{B}(t, t)$, $t = 2, \dots, T$ を推定する．

$$\hat{\mathbf{N}}(t) = \mathbf{B}(t, t)\hat{\mathbf{N}}(t) + \mathbf{e}(t)$$

4. 最後に過去の変数からの影響 $\mathbf{B}(t, t-\tau)$ ($\tau > 0$) を以下により求める．

$$\hat{\mathbf{B}}(t, t-\tau) = (\mathbf{I} - \hat{\mathbf{B}}(t, t))\hat{\mathbf{M}}(t, t-\tau) \quad (\tau > 0)$$

出力として，同時点の変数間の影響 $\hat{\mathbf{B}}(t, t)$ と過去の変数の影響 $\hat{\mathbf{B}}(t, t-\tau)$ ($\tau > 0$) を得る．

6. 評価実験

本実験では人工的に生成したデータを用いて実験を行った．実験は経時 LiNGAM アルゴリズムと AR-LiNGAM アルゴリズムの推定精度の比較を行った．

6.1 実験設定

以下の手順で実験に使用するデータを生成した．

1. まず，同時点における変数間の影響を表す狭義下三角行列 $\mathbf{B}(t, t)$ ($t = 1, \dots, T$) を下三角の成分が全て非ゼロである密の状態になるように設定した．
2. $\mathbf{B}(t, t)$ ($t = 1, \dots, T$) の非ゼロ要素を $[-1.5, -0.5] \cup [0.5, 1.5]$ の範囲からランダムに等確率で選んだ値に置き換えた．また $l = 1$ として 1 時点前からの影響を表す $\mathbf{B}(t, t-1)$ ($t = 2, \dots, T$) の各要素を $[-0.5, 0] \cup [0, 0.5]$ の範囲から等確率でランダムに選んだ．
3. 外乱 $e_i(t)$ の分散を $[1, 2]$ の間から等確率で選び，18 種類の非ガウス分布 [Bach 02] を用いて時点数 T の外乱 $e_i(t)$ ($t = 2, \dots, T$) をそれぞれ独立に生成した． $t = 1$ においてはまず潜在交絡変数 f を $[0, 1]$ の一様分布を用いて生成した．それから $[0, 1]$ のランダムな変数を持つ列ベクトル \mathbf{c} を用いて以下の式から各 $e_i(1)$ が f の影響を受けるように設定した．

$$\mathbf{x}(1) = \mathbf{B}(1, 1)\mathbf{x}(1) + \mathbf{e}(1) + \mathbf{c}f$$

*2 $\mathbf{N}(t)$ は $\mathbf{X}(t)$ と同じサイズの $p \times n$ 行列．

4. 式 (5) に従って観測変数 $x_i(t)$ を生成した。この際に、変数間の因果的順序を推定アルゴリズムに対して隠すために $\mathbf{X}(t)$ の行の順序をランダムに入れ替えた。また、この順序に従って $\mathbf{B}(t, t)$ の行と列を同時に入れ替えた。

生成したデータに対し AR-LiNGAM アルゴリズム及び経時 LiNGAM アルゴリズムを適用して同時点間の因果関係を表す $\mathbf{B}(t, t)$ を推定した。AR-LiNGAM アルゴリズムではランダムに一つの時系列 $\mathbf{x}(1), \dots, \mathbf{x}(T)$ を選択して並べた $p \times T$ 行列 \mathbf{X}_{AR} を入力として出力 \mathbf{B}_T を比較対象とした。そして、それぞれの手法に対し推定を 10 回繰り返して、各 $\mathbf{B}(t, t), \mathbf{B}, \mathbf{B}(t), \mathbf{B}_T$ に対する推定精度を比較した。

6.2 結果

AR-LiNGAM 及び経時 LiNGAM アルゴリズムにおいて推定された $\hat{\mathbf{B}}_0, \hat{\mathbf{B}}(t, t)$ に対して二乗平均平方根誤差 (Root Mean Square Error, RMSE) の時点 $t = 2, \dots, T$ に対する合計を比較した。 $\mathbf{B}(t, t)$ の RMSE は次式で表される。

$$RMSE(\mathbf{B}(t, t)) = \sqrt{\frac{1}{p^2(T-1)} \sum_{t=2}^T \sum_{i,j=1}^p (\mathbf{B}_{true}(t, t)(i, j) - \hat{\mathbf{B}}(t, t)(i, j))^2}$$

$\hat{\mathbf{B}}_0$ に対する RMSE も同様にして計算した。以下、表 1 で $T = 20$ における誤差、表 2 で $T = 50$ における誤差をそれぞれ示す。

表からわかるように経時 LiNGAM アルゴリズムを用いることで因果構造の変化する経時データの推定精度が向上したと言える。

6.3 考察

表からわかるように、経時 LiNGAM アルゴリズムではサンプルサイズを増やしていくと推定精度が向上したが AR-LiNGAM では向上しなかった。これは因果構造が時間とともに変化するという仮定が満たされていないために、正しく推定出来なかったからだと考えられる。また、変数の次元数が増えたとその推定精度により顕著な差が見られた。故に、経時モデ

表 1 $T = 20$ の場合の推定誤差を比較した表。誤差が小さい方を赤字で表した。

		サンプルサイズ			
		100	200	500	1000
AR-LiNGAM	n=2	0.50	0.53	0.51	0.51
アルゴリズム	n=3	0.61	0.60	0.61	0.61
	n=5	0.70	0.69	0.70	0.71
経時 LiNGAM	n=2	0.36	0.35	0.22	0.14
アルゴリズム	n=3	0.41	0.32	0.25	0.16
	n=5	0.44	0.38	0.24	0.20

表 2 $T = 50$ の場合の推定誤差を比較した表。誤差が小さい方を赤字で表した。

		サンプルサイズ			
		100	200	500	1000
AR-LiNGAM	n=2	0.53	0.53	0.53	0.52
アルゴリズム	n=3	0.60	0.59	0.61	0.60
	n=5	0.77	0.71	0.72	0.79
経時 LiNGAM	n=2	0.35	0.33	0.25	0.12
アルゴリズム	n=3	0.39	0.35	0.27	0.18
	n=5	0.66	0.40	0.30	0.17

ル LiNGAM アルゴリズムは因果構造が時間とともに変化する場合にもうまく対応できることが示された。

7. 結論

時間の経過とともに観測した経時データに対して、その構造が時間的に変化する場合や時点間の影響が存在する場合に正しく構造を推定できるような新たなモデル及びその推定アルゴリズムを提案した。そして、その手法が従来手法よりも優れた性能を持つことを実験によって示した。実データにおいては因果構造及び影響の強さが時間的に変化する、または過去の時点の変数の影響を受けるような経時データは少なくない。従って、本研究の手法を用いることで従来よりも高い精度で経時データの因果構造を推定できるようになると期待される。

参考文献

- [Bach 02] Bach, F. R. and Jordan, M. I.: Kernel independent component analysis, *Journal of Machine Learning Research*, Vol. 3, pp. 1–48 (2002)
- [Bollen 89] Bollen, K. A.: *Structural equations with latent variables*, John Wiley (1989)
- [Hoyer 08] Hoyer, P., Shimizu, S., Kerminen, A., and Palviainen, M.: Estimation of causal effects using linear non-Gaussian causal models with hidden variables, *International Journal of Approximate Reasoning*, Vol. 49, No. 2, pp. 362–378 (2008)
- [Hyvärinen 10] Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P.: Estimation of a structural vector autoregression model using non-Gaussianity, *Journal of Machine Learning Research*, Vol. 11, pp. 1709–1731 (2010)
- [Shimizu 06] Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A.: A linear non-Gaussian acyclic model for causal discovery, *Journal of Machine Learning Research*, Vol. 7, pp. 2003–2030 (2006)
- [Shimizu 11] Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P., and Bollen, K.: DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model., *Journal of Machine Learning Research* (2011)
- [Shimizu 12] Shimizu, S.: Joint estimation of linear non-Gaussian acyclic models, *Neurocomputing*, Vol. 81, pp. 104–107 (2012)
- [Swanson 97] Swanson, N. and Granger, C.: Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions, *Journal of the American Statistical Association*, Vol. 92, No. 437, pp. 357–367 (1997)
- [尾崎 98] 尾崎統, 北川源四郎: 時系列解析の方法 (1998)