Instrumental Variable Tests for Directed Acyclic Graph Models

Manabu Kuroki
Department of Systems Innovation
Graduate School of Engineering Science
Osaka University
Machikaneyama-cho, Toyonaka, Osaka
mkuroki@sigmath.es.osaka-u.ac.jp

Zhihong Cai
Department of Biostatistics
Graduate School of Public Health
Kyoto University
Yoshida, Konoe-cho, Sakyo-ku, Kyoto
cai@pbh.med.kyoto-u.ac.jp

Abstract

Consider a case where cause-effect relationships among variables can be described as a directed acyclic graph model. The instrumental variable (IV) method is well known as a powerful tool for inferring the total effect of a treatment variable X on a response variable Y, even if there exist unobserved confounders between X and Y. This paper proposes a criterion to search for covariates which satisfy the IV conditions in linear structural equation models. In addition, it is shown that our criterion is applicable to the case where all variables in a directed acyclic graph are discrete. The results of this paper provide a statistical justification for testing whether the statistical data is generated by a model involving IVs.

1 INTRODUCTION

Estimating total effects is an important problem in many scientific domains. The back door criterion (e.g. Pearl, 2000) and the instrumental variable (IV) method (e.g. Bowden and Turkington, 1984) are two main powerful tools to estimate total effects. However, in the case where unobserved confounders exist, it may be difficult to apply the back door criterion to estimating total effects. Under such circumstances, it is beneficial to utilize the IV method. Here, a total effect τ_{yx} of X on Y is a measure for evaluating causal effects, and can be interpreted as the change of the mean of a response variable Y when a treatment variable X is changed by one unit through an external intervention.

The IV method is well known as a powerful tool to estimate total effects from observational data, in case where unobserved confounders are at presence. Therefore, the IV method has been studied by many researchers in practical sciences (e.g. Angrist et al, 1996; Bowden and Turkington, 1984; Greenland, 2000). It is also one of current hot topics in artificial intelligence and statistics (e.g. Bollen and Bauer, 2004; Brito and Pearl, 2002; Chu et al., 2001; Kuroki and Cai, 2004; Lauritzen, 2004; Pearl, 2004; Scheines et al., 2001). In the IV method, when we consider a linear regression model $Y = \tau_{yx}X + U$ in the case where an unobserved confounder U is correlated with X, the total effect τ_{yx} can be estimated through an observed variable Z (called an IV) which is correlated with X but not with U, and is given by σ_{yz}/σ_{xz}. Here, σ_{yz} and σ_{xz} are a covariance between Y and Z and a covariance between X and Z, respectively.

As many researchers have pointed out, a serious problem of the IV method is that it is difficult to test whether a covariate satisfies the IV conditions from observational data, since U is an unobserved variable. In order to solve this problem, Pearl (1995) provided the instrumental inequality as a necessary condition to test whether a variable Z is an IV relative to (X, Y), in the case where X is a discrete variable. However, there is no testing method for IVs when X is continuous. Considering that there are many applications regarding the IV methods in linear structural equation models, it is necessary to develop a testing criterion in this framework.

In this paper, we first introduce the tetrad difference which has been used to test whether statistical dependencies among observed variables are due to a single common factor in factor analysis. It is shown that the tetrad difference can be applied to linear case as well as discrete case. Although searching for one latent common factor is quite different from searching for an IV, we will show that similar consideration is helpful to provide a criterion to search for covariates which satisfy the IV conditions. Pearl (1995) provided an inequality to search for one instrumental variable, while we propose an equation to search for two or more IVs. Hence, our method may provide a tighter condition than Pearl (1995). Since it is the usual case in practical science that there exist many observed covariates as well as unobserved confounders, if we can find two or more covariates satisfying the IV conditions, it is possible to apply IV method to estimating total effects, whether such covariates are continuous or discrete.

This paper is arranged as follows. In the next section, basic graph terminologies are given. Section 3 describes previous studies about the tetrad difference, and then we show that tetrad difference is applicable to discrete case. On the basis of similar consideration, section 4 provides testing criteria for conditional IVs (Brito and Pearl, 2002) in both continuous case and discrete case. Section 5 illustrates our results with an example. Some further topics are presented in the final section.

2 PRELIMINARIES

2.1 GRAPHS

A directed graph is a pair $G = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ is a finite set of vertices and the set $\boldsymbol{E}$ of arrows is a subset $\boldsymbol{V}\times\boldsymbol{V}$ of ordered pairs of distinct vertices. An arrow pointing from a vertex a to a vertex b indicates $(a, b)\in\boldsymbol{E}$ and $(b, a) \notin \boldsymbol{E}$. The arrow is said to emerge from a or point to b. If there is an arrow pointing from a to b, a is said to be a parent of b, and b a child of a. The set of parents of b is denoted as $pa(b)$, and the set of children of a as $ch(a)$.

A path between a and b is a sequence $a = a_0$, a_1, $\cdots$, $b = a_n$ of distinct vertices such as $(a_{i-1}, a_i)\in\boldsymbol{E}$ or $(a_i, a_{i-1})\in\boldsymbol{E}$ for all $i = 1, 2, \cdots, n$. A directed path from a to b is a sequence $a = a_0$, a_1, $\cdots$, $b = a_n$ of distinct vertices such as $(a_{i-1}, a_i)\in\boldsymbol{E}$ and $(a_i, a_{i-1})\notin\boldsymbol{E}$ for all $i = 1, \cdots, n$. If there exists a directed path from a to b, a is said to be an ancestor of b and b a descendant of a. When the set of descendants of a is denoted as $de(a)$, the vertices in $\boldsymbol{V}\backslash(de(a)\cup\{a\})$ are said to be the nondescendants of a. A directed cycle is a sequence a_0, a_1, $\cdots$, a_{n-1} of distinct vertices such as (1) $(a_{i-1}, a_i)\in\boldsymbol{E}$ and $(a_i, a_{i-1}) \notin\boldsymbol{E}$ for all $i = 1, \cdots, n-1$, and (2) $(a_{n-1}, a_0)\in\boldsymbol{E}$.

If a directed graph has no directed cycles, then the graph is said to be a directed acyclic graph.

2.2 BAYESIAN NETWORKS

Let $\mathrm{p}_{v_1,\ldots,v_n}$ be a strictly positive joint distribution of $\boldsymbol{V} = \{V_1, \cdots, V_n\}$ and $\mathrm{p}_{v_i|v_j}$ the conditional distribution of V_i given V_j. Similar notations are used for other distributions. Then, a graphical representation of the conditional independencies among $V_1, \cdots, V_n$ in form of a directed acyclic graph is given in the following way:

DEFINITION 1 (BAYESIAN NETWORK)
Suppose that a set of variables $\boldsymbol{V} = \{V_1, \ldots, V_n\}$ and a directed acyclic graph $G = (\boldsymbol{V}, \boldsymbol{E})$ are given. When the joint distribution of $\boldsymbol{V}$ is factorized recursively according to the graph G as the following equation, the graph is called a Bayesian network.

$$\mathrm{p}_{v_1,\cdots,v_n} = \prod_{i=1}^{n} \mathrm{p}_{v_i|pa(v_i)}, \tag{1}$$

When $pa(V_i)$ is an empty set, $\mathrm{p}_{v_i|pa(v_i)}$ is the marginal distribution of v_i. □

If a joint distribution is factorized recursively according to the graph G, the conditional independencies implied by the factorization (1) can be obtained from the graph G according to the following criterion (Pearl, 1988):

DEFINITION 2 (D-SEPARATION)
Let $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ be three disjoint subsets of vertices in a Bayesian network G. Then $\boldsymbol{Z}$ is said to d-separate $\boldsymbol{X}$ from $\boldsymbol{Y}$, if along every path between a vertex in $\boldsymbol{X}$ and a vertex in $\boldsymbol{Y}$ there exists a vertex w which satisfies one of the following three conditions:

1. w is in $\boldsymbol{Z}$, and one arrow on the path points to w and the other arrow on the path emerges from w.
2. w is in $\boldsymbol{Z}$, and two arrows on the path emerge from w.
3. Neither w nor any descendant of w is in $\boldsymbol{Z}$, but two arrows on the path point to w.

If a path satisfies one of the conditions above, the path is said to be blocked by $\boldsymbol{Z}$. □

It can be shown that if $\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$ in a Bayesian network G then $\boldsymbol{X}$ is conditionally independent of $\boldsymbol{Y}$ given $\boldsymbol{Z}$ in the corresponding recursive factorization (1) (Geiger et al., 1990).

2.3 LINEAR STRUCTURAL EQUATION MODEL

In the case where variables in $\boldsymbol{V}$ are normally distributed, equation (1) provides a linear structural equation model

$$V_i = \sum_{V_j\in pa(V_i)} \alpha_{v_iv_j} V_j + \epsilon_{v_i} \qquad i = 1, \ldots, n, \tag{2}$$

where $\alpha_{v_iv_j}\,(\neq 0)$ is called a path coefficient. In the linear structural equation models discussed in this paper, it is assumed that each variable in $\boldsymbol{V}$ has mean 0, and that random disturbances $\epsilon_{v_1}, \cdots, \epsilon_{v_n}$ are independent and normally distributed. For further details of linear structural equation models, see Bollen (1989).

Let $\sigma_{xy\cdot z}$ and $\beta_{yx\cdot z}$ be the conditional covariance of X and Y given $\boldsymbol{Z}$ and the regression coefficient of x in the regression model of Y on x and $\boldsymbol{z}$, respectively. $\rho_{xy\cdot z}$ is a partial correlation coefficient of X and Y given $\boldsymbol{Z}$, and $\sigma_{yy\cdot z}$ is a conditional variance of Y given $\boldsymbol{Z}$. In the case where $\boldsymbol{Z}$ is an empty set, σ_{xy} and β_{yx} are the covariance of X and Y and the regression coefficient of x in the regression model of Y on x, respectively. In addition, ρ_{xy} and σ_{yy} are the correlation coefficient of X and Y and the variance of Y, respectively. Similar notations are used for other parameters.

When a linear structural equation model is given according to the graph G, if $\boldsymbol{Z}$ d-separates X from Y, then $\sigma_{xy\cdot z} = \beta_{yx\cdot z} = 0$ (e.g. Spirtes et al., 1993).

3 TETRAD DIFFERENCE

3.1 LINEAR CASE

We consider the directed acyclic graph shown in Fig.1 and the corresponding linear structural equation model.

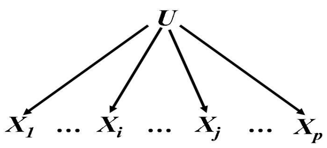

Fig.1: Directed acyclic graph (1)

As shown in Fig.1, U is a parent of each element of $\boldsymbol{X} = (X_1, \cdots, X_p)$ and d-separates any two elements of $\boldsymbol{X}$. The covariance structure corresponding to Fig.1 can be described as

$$\Sigma_{xx} = \Sigma_{xx\cdot u} + \frac{1}{\sigma_{uu}}\Sigma_{xu}\Sigma'_{xu}, \tag{3}$$

where Σ_{xx} and $\Sigma_{xx\cdot u}$ is a covariance matrix of $\boldsymbol{X}$ and a conditional covariance matrix of $\boldsymbol{X}$ given U, respectively. In addition, Σ_{xu} is a covariance matrix between $\boldsymbol{X}$ and U. Similar notations are used for other matrices.

Then, it can be understood that

1) $\Sigma_{xx\cdot u}$ is a positive diagonal matrix which satisfies $\mathrm{rank}(\Sigma_{xx} - \Sigma_{xx\cdot u}) = 1$ and $\Sigma_{xx} - \Sigma_{xx\cdot u}$ is a positive semidefinite matrix, and

2) Since $\sigma_{x_ix_j}\sigma_{x_kx_l} = (\sigma_{x_iu}\sigma_{x_ju}/\sigma_{uu})(\sigma_{x_ku}\sigma_{x_lu}/\sigma_{uu}) = \sigma_{x_ix_k}\sigma_{x_jx_l}$,

$$\sigma_{x_ix_j}\sigma_{x_kx_l} - \sigma_{x_ix_k}\sigma_{x_jx_l} = 0 \tag{4}$$

$$\sigma_{x_ix_i} - \frac{\sigma_{x_ix_j}\sigma_{x_ix_k}}{\sigma_{x_jx_k}} = \sigma_{x_ix_i} - \frac{\sigma^2_{x_iu}}{\sigma_{uu}} > 0$$

can be obtained.

For any distinct variables $X_i, X_j, X_k, X_l \in \boldsymbol{X}(\subset \boldsymbol{V})$, Spearman (1904, 1928) defined the tetrad difference as the left hand side of equation (4). When the tetrad difference is equal to zero, it is called a vanishing tetrad difference. It is stated that the vanishing tetrad difference implies that the observed statistical dependencies can be well explained by a single factor model $X_i = \lambda_i U + \epsilon_{x_i}$ $(i = 1, \cdots, p)$, where λ_i is a constant value and $\epsilon_{x_1}, \cdots, \epsilon_{x_p}$ are independent.

The tetrad difference has been studied by many researchers for decades. Bekker and de Leeuw (1987) summarized the previous results into a single comprehensive theorem. To present this theorem, the following preliminaries are required. When there exists such a positive semidefinite diagonal matrix Ω that $\Sigma_{xx} - \Omega$ is a positive semidefinite matrix and $\mathrm{rank}(\Sigma_{xx} - \Omega) = 1$, Σ_{xx} is said to be a Spearman matrix, which provides statistical evidence that the observed statistical dependencies can be well explained by a single factor model (Bekker and de Leeuw, 1987). In addition, any element of Σ_{xx} is assumed to be nonzero. With this preparation, Bekker and de Leeuw (1987) provided the following theorem:

THEOREM 1 (Bekker and de Leeuw, 1987) A covariance matrix $\Sigma_{xx} = (\sigma_{x_ix_j})$ of a set $\boldsymbol{X}$ including four or more variables is a Spearman matrix if, and only if, after sign changes of rows and corresponding columns, all its elements are positive and such that

$$\sigma_{x_ix_j}\sigma_{x_kx_l} - \sigma_{x_ix_k}\sigma_{x_jx_l} = 0 \tag{5}$$

and

$$\frac{\sigma_{x_ix_j}\sigma_{x_ix_k}}{\sigma_{x_jx_k}} \leq \sigma_{x_ix_i} \tag{6}$$

for any distinct variables $X_i, X_j, X_k, X_l \in \boldsymbol{X}$. □

If Theorem 1 holds true, then we can justify the assumption that the observed statistical dependencies can be generated by a single common factor. Regarding the further discussion of models implied by Theorem 1, see Bollen and Ting (2000), Spirtes et al. (1993) and Xu and Pearl (1987).

3.2 DISCRETE CASE

In this section, we show that the similar result of Theorem 1 holds true in the case where all variables in a directed acyclic graph are dichotomous variables.

For dichotomous variables $X, Y \in \boldsymbol{V}$ $(a, b \in \{0, 1\})$, let $\mathrm{p}_{X=a,Y=b}$ be a joint probability of $X = a$ and $Y = b$, and $\mathrm{p}_{X=a}$ a marginal probability of $X = a$. In addition, $\mathrm{p}_{X=a|Y=b}$ is a conditional probability of $X = a$ given $Y = b$ $(a, b \in \{0, 1\})$. Similar notations are used for other parameters.

Letting $\sigma_{x_i x_j} = \mathrm{p}_{x_i=1,x_j=1} - \mathrm{p}_{x_i=1}\mathrm{p}_{x_j=1}$ and $\sigma_{x_i x_i} = \mathrm{p}_{x_i=1}(1-\mathrm{p}_{x_i=1})$ (These are correspondent to a covariance between X_i and X_j and a variance of X_i, respectively), based on the Bayesian network shown in Fig.1, since

$$\begin{aligned}\sigma_{x_i x_j} &= (\mathrm{p}_{x_i=1|u=0} - \mathrm{p}_{x_i=1|u=1}) \\ &\quad \times(\mathrm{p}_{x_j=1|u=0} - \mathrm{p}_{x_j=1|u=1})\mathrm{p}_{u=0}\mathrm{p}_{u=1}\end{aligned}$$

for any X_i and X_j, we can obtain

$$\begin{aligned}\Sigma_{xx} &= \mathrm{Diag}[\omega_{11}, \cdots, \omega_{pp}] \\ &\quad +\mathrm{p}_{u=0}\mathrm{p}_{u=1}\begin{pmatrix} \mathrm{p}_{x_1=1|u=0} - \mathrm{p}_{x_1=1|u=1} \\ \vdots \\ \mathrm{p}_{x_p=1|u=0} - \mathrm{p}_{x_p=1|u=1} \end{pmatrix} \\ &\quad \times(\mathrm{p}_{x_1=1|u=0} - \mathrm{p}_{x_1=1|u=1}, \cdots, \mathrm{p}_{x_p=1|u=0} - \mathrm{p}_{x_p=1|u=1}).\end{aligned}$$

Here,

$$\omega_{ii} = \sigma_{x_i x_i} - (\mathrm{p}_{x_i=1|u=0} - \mathrm{p}_{x_i=1|u=1})^2\mathrm{p}_{u=0}\mathrm{p}_{u=1}$$

for $i = 1, 2, \cdots, p$. Then, since

$$\begin{aligned}\omega_{ii} &= \mathrm{p}_{u=0}(\mathrm{p}_{x_i=1|u=0} - \mathrm{p}^2_{x_i=1|u=0}) \\ &\quad +\mathrm{p}_{u=1}(\mathrm{p}_{x_i=1|u=1} - \mathrm{p}^2_{x_i=1|u=1})\geq 0\end{aligned}$$

for any X_i, we can understand that Σ_{xx} can be also expressed as the form of $\Omega + \boldsymbol{\lambda}\boldsymbol{\lambda}'$, where Ω is a positive definite diagonal matrix and $\boldsymbol{\lambda}$ is a vector.

Thus, based on the consideration above, it is shown that the same result as Theorem 1 holds true in the case where all variables in a directed acyclic graph are dichotomous. Regarding the further discussion, see Pearl and Tarci (1986).

4 INSTRUMENTAL VARIABLE (IV) METHOD

4.1 IDENTIFICATION

In this section, we introduce the conditional instrumental variable (IV) method (Brito and Pearl, 2002) as the identifiability criterion for total effects. Here, a total effect τ_{yx} of X on Y is defined as the total sum of the products of the path coefficients on the sequence of arrows along all directed paths from X to Y. When a total effect can be determined uniquely from the correlation parameters of observed variables, it is said to be identifiable, that is, it can be estimated consistently. Let $G_{\underline{X}}$ be the graph obtained by deleting from a graph G all arrows emerging from vertices in $\boldsymbol{X}$.

DEFINITION 3 (CONDITIONAL IV)
If a set $\boldsymbol{W} \cup \{Z\}$ of variables satisfies the following conditions relative to an ordered pair (X, Y) of variables in a directed acyclic graph G, then Z is said to be a conditional instrumental variable (IV) given $\boldsymbol{W}$ relative to (X, Y).

1. A set $\boldsymbol{W}$ of variables is a subset of nondescendants of X and Y in G, and
2. $\boldsymbol{W}$ d-separates Z from Y but not from X in $G_{\underline{X}}$.

□

In linear structural equation models, if an observed variable Z is a conditional IV given a set $\boldsymbol{W}$ of observed variables relative to (X, Y), then the total effect τ_{yx} of X on Y is identifiable, and is given by $\sigma_{yz\cdot w}/\sigma_{xz\cdot w}$ (Brito and Pearl, 2002). When $\boldsymbol{W}$ is an empty set in Definition 3, Z is called an instrumental variable (IV) (Bowden and Turkington, 1984).

When the sample conditional covariances $\hat{\sigma}_{yz\cdot w}$ and $\hat{\sigma}_{xz\cdot w}$ of the conditional covariances $\sigma_{yz\cdot w}$ and $\sigma_{xz\cdot w}$ are used to estimate the total effect τ_{yx}, by the delta method (Anderson, 1986), the asymptotic variance of $\hat{\sigma}_{yz\cdot w}/\hat{\sigma}_{xz\cdot w}$ is given by

$$a.var\left(\frac{\hat{\sigma}_{yz\cdot w}}{\hat{\sigma}_{xz\cdot w}}\right) = \frac{\sigma_{yy\cdot w}/\sigma_{xx\cdot w} - 2\beta_{yx\cdot w}\tau_{yx} + \tau^2_{yx}}{n\rho^2_{xz\cdot w}} \quad (7)$$

(e.g. Kuroki and Cai, 2004), where n is the sample size, and $\rho_{xz\cdot w}$ is a partial correlation coefficient of X and Z given $\boldsymbol{W}$. Clearly, for a given $\boldsymbol{W}$, the smaller the value of the $\rho^2_{xz\cdot w}$ is, the larger the asymptotic variance of $\hat{\sigma}_{yz\cdot w}/\hat{\sigma}_{xz\cdot w}$ is. This fact provides an IV selection criterion for estimating total effects (e.g. Kuroki and Cai, 2004).

4.2 LINEAR CASE

In this section, we propose a criterion for testing whether statistical data is generated by a model involving IVs. For this purpose, consider the directed acyclic graph shown in Fig.2.

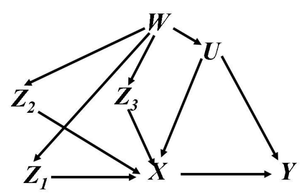

Fig.2: Directed acyclic graph (2)

Fig.2 shows that Z_1, Z_2 and Z_3 are conditional IVs given $\boldsymbol{W}$ relative to (X, Y). In addition, $Z_i \perp\!\!\!\perp Z_j|\boldsymbol{W}$ holds true $(i, j = 1, 2, 3(i \neq j))$. Then, the following conditional covariances can be obtained:

$$\begin{aligned}\sigma_{z_i z_j\cdot xw} &= -\frac{\sigma_{z_i x\cdot w}\sigma_{z_j x\cdot w}}{\sigma_{xx\cdot w}} && (i, j = 1, 2, 3(i \neq j)) \\ \sigma_{z_i y\cdot xw} &= \sigma_{z_i y\cdot w} - \frac{\sigma_{z_i x\cdot w}\sigma_{yx\cdot w}}{\sigma_{xx\cdot w}} && (i = 1, 2, 3).\end{aligned}$$

Here, by $\tau_{yx} = \sigma_{yz_i \cdot w}/\sigma_{xz_i \cdot w} (i = 1, 2, 3)$, we can obtained

$$\sigma_{z_i y \cdot xw} = \sigma_{xz_i \cdot w}(\tau_{yx} - \beta_{yx \cdot w}).$$

Hence, letting $\boldsymbol{Z} = \{Z_1, Z_2, Z_3\}$, the conditional covariance matrix of $\boldsymbol{S} = \boldsymbol{Z} \cup \{Y\}$ given X and $\boldsymbol{W}$ can be provided as

$$\begin{aligned}
&\Sigma_{ss \cdot xw} \\
&= \begin{pmatrix} \Sigma_{zz \cdot w} & \mathbf{0} \\ \mathbf{0}' & \sigma_{yy \cdot xw} + \sigma_{xx \cdot w}(\tau_{yx} - \beta_{yx \cdot w})^2 \end{pmatrix} \\
&\quad - \begin{pmatrix} \dfrac{\Sigma_{zx \cdot w}}{\sqrt{\sigma_{xx \cdot w}}} \\ -\sqrt{\sigma_{xx \cdot w}}(\tau_{yx} - \beta_{yx \cdot w}) \end{pmatrix} \\
&\quad \times \left(\frac{\Sigma'_{zx \cdot w}}{\sqrt{\sigma_{xx \cdot w}}}, -\sqrt{\sigma_{xx \cdot w}}(\tau_{yx} - \beta_{yx \cdot w})\right) \\
&= \Omega - \boldsymbol{\lambda}\boldsymbol{\lambda}', \qquad (8)
\end{aligned}$$

where $\Sigma'_{xz \cdot w}$ is a conditional covariance vector of X and $\boldsymbol{Z}$ given $\boldsymbol{W}$. Similar notations are used for other vectors. That is, a necessary condition that Z_1, Z_2 and Z_3 are conditional IVs given $\boldsymbol{W}$ relative to (X, Y) is that $\Sigma_{ss \cdot xw}$ can be reformulated as equation (8) through a positive diagonal matrix Ω and a vector $\boldsymbol{\lambda}$. This necessary condition, which is testable by observational data, holds true when there exist two or more covariates which satisfy the IV conditions. In addition, from equation (8), it can be understood that

1) $\sigma_{z_1 z_2 \cdot xw}\sigma_{z_3 y \cdot xw} = \sigma_{z_1 z_3 \cdot xw}\sigma_{z_2 y \cdot xw} = \sigma_{z_2 z_3 \cdot xw}\sigma_{z_1 y \cdot xw}$ hold true in Fig.2, which is similar to the vanishing tetrad differences.

2) when the (i, j) component of $\Sigma_{ss \cdot xw}$ in equation (8) is positive, letting Q be the matrix obtained from the unit matrix by multiplying the i th or j th diagonal component with -1, we can obtain

$$Q\Sigma_{ss \cdot xw}Q' = Q\Omega Q' - (Q\boldsymbol{\lambda})(Q\boldsymbol{\lambda})'.$$

That is, by sign changes of rows and corresponding columns, all the off-diagonal elements of $\Sigma_{ss \cdot xw}$ in equation (8) can become negative, since multiplications of the column vector $\boldsymbol{\lambda}$ with Q finally turn all the elements of $Q\boldsymbol{\lambda}$ into positive.

3) $\Sigma_{ss \cdot xw} - \Omega$ is a negative semidefinite matrix and $\text{rank}(\Sigma_{ss \cdot xw} - \Omega) = 1$.

4) the (i, i) component of Ω is equal to the conditional variance of Z_i given $\boldsymbol{W}$ $(i = 1, 2, 3)$ but the $(4, 4)$ component of Ω is not equal to the conditional variance of Y given $\boldsymbol{W}$.

5) When $\tau_{yx} = \beta_{yx \cdot w}$ holds true, the last component corresponding to Y in $\boldsymbol{\lambda}$ is equal to zero. This occurs in the case where $\boldsymbol{W}$ satisfies "the back door criterion" relative to (X, Y) (for "the back door criterion", refer to Pearl (2000)).

On the basis of the considerations above, when any element of $\Sigma_{ss \cdot xw}$ is assumed to be nonzero and $\beta_{yx \cdot w}$ is not consistent with τ_{yx}, the following theorem can be obtained:

THEOREM 2

For a covariance matrix $\Sigma_{ss \cdot xw}$ of a set $\boldsymbol{S} = \{Z_1, \cdots, Z_p\}$ (here, letting $Z_p = Y$) including four or more variables, after sign changes of rows and corresponding columns of $\Sigma_{ss \cdot xw}$, all its off-diagonal elements are negative and such that

$$\sigma_{z_i z_j \cdot xw}\sigma_{z_k z_l \cdot xw} = \sigma_{z_i z_k \cdot xw}\sigma_{z_j z_l \cdot xw} \qquad (9)$$

and

$$\frac{\sigma_{z_i z_k \cdot xw}\sigma_{z_i z_j \cdot xw}}{\sigma_{z_j z_k \cdot xw}} \leq \sigma_{z_i z_i \cdot xw} \qquad (10)$$

for any distinct variables $Z_i, Z_j, Z_k, Z_l \in \boldsymbol{S}$ if, and only if, there exist a positive definite diagonal matrix Ω and a vector $\boldsymbol{\lambda}$ satisfying that

$$\Sigma_{ss \cdot xw} = \Omega - \boldsymbol{\lambda}\boldsymbol{\lambda}'. \qquad (11)$$

□

PROOF OF THEOREM 2

First, suppose that there exist a positive definite diagonal matrix Ω and a vector $\boldsymbol{\lambda}$, which satisfies equation (11). Then, from the consideration 2) stated above, it is trivial that all its off-diagonal elements can be negative after sign changes of rows and corresponding columns of $\Sigma_{ss \cdot xw}$. In addition, since $\sigma_{z_i z_j \cdot xw} = -\lambda_i \lambda_j$ holds true for any Z_i and Z_j, we can obtain

$$\begin{aligned}
&\sigma_{z_i z_j \cdot xw}\sigma_{z_k z_l \cdot xw} = (-\lambda_i\lambda_j)(-\lambda_k\lambda_l) \\
&= \sigma_{z_i z_k \cdot xw}\sigma_{z_l z_j \cdot xw} = \sigma_{z_i z_l \cdot xw}\sigma_{z_k z_j \cdot xw}
\end{aligned}$$

and

$$\begin{aligned}
&\sigma_{z_i z_i \cdot xw} - \frac{\sigma_{z_i z_j \cdot xw}\sigma_{z_i z_k \cdot xw}}{\sigma_{z_j z_k \cdot xw}} \\
&= \sigma_{z_i z_i \cdot xw} - \frac{(-\lambda_i\lambda_j)(-\lambda_i\lambda_k)}{(-\lambda_j\lambda_k)} \\
&= \sigma_{z_i z_i \cdot xw} + \lambda_i^2 > 0.
\end{aligned}$$

Thus, equations (9) and (10) can be obtained from $\Sigma_{ss \cdot xw}$.

Next, as we can assume that all off-diagonal elements of $\Sigma_{ss \cdot xw}$ are negative, by sign changes of rows and corresponding columns of $\Sigma_{ss \cdot xw}$, it will be shown that a p dimensional vector $\boldsymbol{\lambda}' = (\lambda_1, \cdots, \lambda_p)$ exists such that $\Sigma_{ss \cdot xw} + \boldsymbol{\lambda}\boldsymbol{\lambda}'$ is a positive diagonal matrix. Each element of $\sigma_{z_i z_j \cdot xw}$ $(Z_i \neq Z_j)$ can be described as

$$\begin{aligned}
\sigma_{z_i z_j \cdot xw} &= \frac{\sigma_{z_i z_k \cdot xw}\sigma_{z_j z_l \cdot xw}}{\sigma_{z_k z_l \cdot xw}} \\
&= -\left(\left|\frac{\sigma_{z_i z_k \cdot xw}\sigma_{z_i z_j \cdot xw}}{\sigma_{z_j z_k \cdot xw}}\right|\right)^{1/2}\left(\left|\frac{\sigma_{z_j z_k \cdot xw}\sigma_{z_i z_j \cdot xw}}{\sigma_{z_i z_k \cdot xw}}\right|\right)^{1/2} \\
&= -\lambda_i\lambda_j.
\end{aligned}$$

Here, from equation (9), it is noted that $\lambda_i = (|\sigma_{z_iz_k\cdot xw}\sigma_{z_iz_j\cdot xw}/\ \sigma_{z_jz_k\cdot xw}|)^{1/2}$ takes the same value regardless of the choice of Z_j and Z_k, since

$$\frac{\sigma_{z_iz_k\cdot xw}\sigma_{z_iz_j\cdot xw}}{\sigma_{z_jz_k\cdot xw}} = \frac{\sigma_{z_iz_l\cdot xw}\sigma_{z_iz_j\cdot xw}}{\sigma_{z_jz_l\cdot xw}} = \frac{\sigma_{z_iz_l\cdot xw}\sigma_{z_iz_m\cdot xw}}{\sigma_{z_mz_l\cdot xw}},$$

$(Z_i, Z_j, Z_k, Z_l$ and Z_m are distinct). On the other hand, noting that equation (10) is automatically satisfied since all off-diagonal elements of $\Sigma_{ss\cdot xw}$ are negative,

$$\begin{aligned}\sigma_{z_iz_i\cdot xw} &= (\sigma_{z_iz_i\cdot xw} + \left|\frac{\sigma_{z_iz_k\cdot xw}\sigma_{z_iz_j\cdot xw}}{\sigma_{z_jz_k\cdot xw}}\right|) \\ &\quad - \left|\frac{\sigma_{z_iz_k\cdot xw}\sigma_{z_iz_j\cdot xw}}{\sigma_{z_jz_k\cdot xw}}\right| \\ &= (\sigma_{z_iz_i\cdot xw} + \lambda_i^2) - \lambda_i^2.\end{aligned}$$

Therefore, the covariance structure of $\Sigma_{ss\cdot xw}$ can be described as $\Omega = \Sigma_{ss\cdot xw} + \boldsymbol{\lambda}\boldsymbol{\lambda}'$. Here, Ω is a positive diagonal matrix. Q.E.D.

4.3 DISCRETE CASE

In this section, we show that the similar result of Theorem 2 holds true in the case where all variables in a directed acyclic graph are dichotomous variables. For this purpose, based on the directed acyclic graph shown in Fig.2, suppose that $\mathrm{p}_{y|x=1,u=a,w=b} - \mathrm{p}_{y|x=0,u=a,w=b} = \tau_{yx}$ for any value $\boldsymbol{a}\cup\boldsymbol{b}$ in a set $\boldsymbol{U}\cup\boldsymbol{W}$, that is, τ_{yx} is constant regardless of $\boldsymbol{a}\cup\boldsymbol{b}$. Here, $\boldsymbol{U}$ represents a set of unobserved confounders, and $\boldsymbol{W}$ represents a set of observed covariates. In addition, let $\sigma_{xy\cdot w} = \mathrm{p}_{x=1,y=1|w=b} - \mathrm{p}_{x=1|w=b}\mathrm{p}_{y=1|w=b}$ for any $\boldsymbol{b}$. Similar notations are used for other parameters. Then, for Z_1, Z_2, Z_3, it is trivial that $\sigma_{z_iz_j\cdot w} = 0$ holds true from Fig.2. In addition,

$$\begin{aligned}&\mathrm{p}_{y=1|z_i=1,w=b} - \mathrm{p}_{y=1|z_i=0,w=b} \\ &= \sum_a(\mathrm{p}_{y=1|x=1,u=a,w=b} - \mathrm{p}_{y=1|x=0,u=a,w=b}) \\ &\quad \times(\mathrm{p}_{x=1|z_i=1,u=a,w=b} - \mathrm{p}_{x=1|z_i=0,u=a,w=b})\mathrm{p}_{u|w=b} \\ &= \tau_{yx}(\mathrm{p}_{x=1|z_i=1,w=b} - \mathrm{p}_{x=1|z_i=0,w=b}). \qquad (12)\end{aligned}$$

By using $\mathrm{p}_{x=1|w=b} = \mathrm{p}_{z_i=0|w=b}\mathrm{p}_{x=1|z_i=0,w=b} + \mathrm{p}_{z_i=1|w=b}\ \mathrm{p}_{x=1|z_i=1,w=b}$,

$$\begin{aligned}\sigma_{xz_i\cdot w} &= (\mathrm{p}_{x=1|z_i=1,w=b} - \mathrm{p}_{x=1|w=b})\mathrm{p}_{z_i=1|w=b} \\ &= (\mathrm{p}_{x=1|z_i=1,w=b} - \mathrm{p}_{x=1|z_i=0,w=b}) \\ &\quad \times\mathrm{p}_{z_i=0|w=b}\mathrm{p}_{z_i=1|w=b} \qquad (13)\end{aligned}$$

Furthermore, by using $\mathrm{p}_{y=1|w=b} = \mathrm{p}_{y=1|z_i=0,w=b} \times\mathrm{p}_{z_i=0|w=b} + \mathrm{p}_{y=1|z_i=1,w=b}\mathrm{p}_{z_i=1|w=b}$ and equations (12) and (13),

$$\begin{aligned}\sigma_{yz_i\cdot w} &= (\mathrm{p}_{y=1|z_i=1,w=b} - \mathrm{p}_{x|w=b})\mathrm{p}_{z_i=1|w=b} \\ &= (\mathrm{p}_{y|z_i=1,w=b} - \mathrm{p}_{y=1|z_i=0,w=b})\mathrm{p}_{z_i=0|w=b}\mathrm{p}_{z_i=1|w=b} \\ &= \tau_{yx}\sigma_{xz_i\cdot w}.\end{aligned}$$

Therefore, letting

$$\begin{aligned}\Sigma_{ss\cdot xw} &= \begin{pmatrix}\Sigma_{zz\cdot w} & \Sigma_{zy\cdot w} \\ \Sigma'_{zy\cdot w} & \sigma_{yy\cdot w}\end{pmatrix} \\ &\quad -\frac{1}{\sigma_{xx\cdot w}}\begin{pmatrix}\Sigma_{zx\cdot w} \\ \sigma_{yx\cdot w}\end{pmatrix}(\Sigma'_{zx\cdot w}, \sigma_{yx\cdot w}),\end{aligned}$$

we can obtain the same formulation as equation (8) by setting

$$\Omega = \begin{pmatrix}\Sigma_{zz\cdot w} & \mathbf{0} \\ \mathbf{0}' & \sigma_{yy\cdot w} - \dfrac{\sigma^2_{xy\cdot w}}{\sigma_{xx\cdot w}} + \sigma_{xx\cdot w}(\tau_{yx} - \dfrac{\sigma_{xy\cdot w}}{\sigma_{xx\cdot w}})^2\end{pmatrix}$$

and

$$\boldsymbol{\lambda} = \begin{pmatrix}\dfrac{\Sigma_{zx\cdot w}}{\sqrt{\sigma_{xx\cdot w}}} \\ -\sqrt{\sigma_{xx\cdot w}}(\tau_{yx} - \dfrac{\sigma_{xy\cdot w}}{\sigma_{xx\cdot w}})\end{pmatrix}.$$

Thus, based on the consideration above, it can be understood that the same result as Theorem 2 holds true in the case where all variables in a directed acyclic graph are dichotomous.

The instrumental inequality (Pearl, 1995) is well known as a necessary condition for testing whether a variable Z is an IV relative to (X, Y) in case where any element of $\boldsymbol{V}$ is discrete in a Bayesian network. This paper provides another testing criterion in discrete case in addition to Pearl's instrumental inequality.

5 APPLICATION

The above results are applicable to analyze the data obtained from a study on setting up painting conditions of car bodies, reported by Okuno et al. (1986). The data was collected with the purpose of setting up the process conditions, in order to increase transfer efficiency. The size of the sample is 38 and the variables of interest, each of which has zero mean and variance one, are the following:

Painting Condition : Dilution Ratio (X_1), Degree of Viscosity (X_2), Painting Temperature (X_8)

Spraying Condition : Gun Speed (X_3), Spray Distance (X_4), Atomizing Air Pressure (X_5), Pattern Width (X_6), Fluid Output (X_7)

Environment Condition : Temperature (X_9), Degree of Moisture (X_{10})

Response: Transfer Efficiency (Y)

Concerning this process, Kuroki et al. (2003) presented the directed acyclic graph shown in Fig.3.

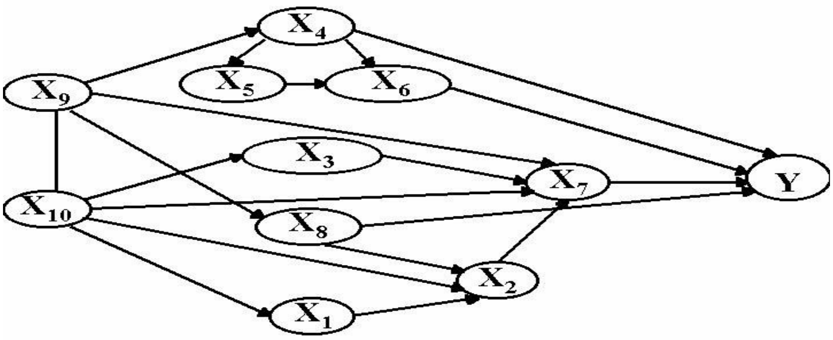

Fig.3 : Directed acyclic graph (3) (Kuroki et al., 2003)

Table 1 : The estimated correlation matrix based on Fig.3 (Kuroki et al., 2003)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y
X_1	1.000	-0.736	-0.152	0.148	0.028	-0.042	0.324	0.216	0.283	-0.496	-0.091
X_2	-0.736	1.000	0.210	-0.331	-0.063	0.095	-0.479	-0.684	-0.635	0.684	0.326
X_3	-0.152	0.210	1.000	-0.091	-0.017	0.026	0.195	-0.134	-0.175	0.307	0.134
X_4	0.148	-0.331	-0.091	1.000	0.191	-0.286	0.184	0.397	0.521	-0.298	-0.614
X_5	0.028	-0.063	-0.017	0.191	1.000	0.291	0.035	0.076	0.099	-0.057	-0.277
X_6	-0.042	0.095	0.026	-0.286	0.291	1.000	-0.053	-0.114	-0.149	0.085	-0.250
X_7	0.324	-0.479	0.195	0.184	0.035	-0.053	1.000	0.396	0.353	-0.146	-0.044
X_8	0.216	-0.684	-0.134	0.397	0.076	-0.114	0.396	1.000	0.761	-0.435	-0.493
X_9	0.283	-0.635	-0.175	0.521	0.099	-0.149	0.353	0.761	1.000	-0.571	-0.475
X_{10}	-0.496	0.684	0.307	-0.298	-0.057	0.085	-0.146	-0.435	-0.571	1.000	0.283
Y	-0.091	0.326	0.134	-0.614	-0.277	-0.250	-0.044	-0.493	-0.475	0.283	1.000

Based on the directed acyclic graph, Kuroki et al. (2003) presented the estimated correlation matrix shown in Table 1.

In order to estimate the effect of X_7 on Y, since both X_1 and X_3 are conditional IVs given $\{X_9, X_{10}\}$, τ_{yx_7} is identifiable through the observation of $\{X_1, X_7, X_9, X_{10}, Y\}$ or $\{X_3, X_7, X_9, X_{10}, Y\}$, and is given by $\tau_{yx_7} = \sigma_{yx_i \cdot x_9 x_{10}}/\sigma_{x_7 x_i \cdot x_9 x_{10}} = 0.195$ $(i = 1, 3)$. In addition, $X_1 \perp\!\!\!\perp X_3 | \{X_9, X_{10}\}$ holds true.

Here, letting $\boldsymbol{S} = \{X_1, X_3, Y\}$, from Table 1, we can obtain

$$\Sigma_{ss \cdot x_7 x_9 x_{10}} = \begin{pmatrix} 0.682 & -0.069 & 0.0134 \\ -0.069 & 0.840 & 0.0130 \\ 0.0134 & 0.0130 & 0.756 \end{pmatrix}$$

$$= \begin{pmatrix} 0.754 & 0.000 & 0.000 \\ 0.000 & 0.906 & 0.000 \\ 0.000 & 0.000 & 0.756 \end{pmatrix}$$

$$- \begin{pmatrix} 0.268 \\ 0.257 \\ -0.051 \end{pmatrix} (0.268, 0.257, -0.051), \quad (14)$$

which is consistent with the form of equation (8). In addition, all the off-diagonal elements of $\Sigma_{ss \cdot x_7 x_9 x_{10}}$ are negative after sign changes of the third rows and the third columns. Furthermore, we can obtain $\sigma_{x_1 x_1 \cdot x_9 x_{10}} = 0.754$ and $\sigma_{x_3 x_3 \cdot x_9 x_{10}} = 0.906$ from Table 1, which are consistent with the (1,1) and the (2,2) components of diagonal matrix in equation (14), respectively. However, the (3,3) component is not equal to $\sigma_{yy \cdot x_9 x_{10}}$. This result indicates that both X_1 and X_3 can be regarded as conditional IVs given $\{X_9, X_{10}\}$ relative to (X_7, Y).

Here, noting that the numerator of equation (7) is not dependent on the selection of the IVs, we can obtain from Table 1,

$$\frac{a.var\left(\dfrac{\hat{\sigma}_{yx_1 \cdot x_9 x_{10}}}{\hat{\sigma}_{x_1 x_7 \cdot x_9 x_{10}}}\right)}{a.var\left(\dfrac{\hat{\sigma}_{yx_3 \cdot x_9 x_{10}}}{\hat{\sigma}_{x_3 x_7 \cdot x_9 x_{10}}}\right)} = \frac{\rho^2_{x_3 x_7 \cdot x_9 x_{10}}}{\rho^2_{x_1 x_7 \cdot x_9 x_{10}}} = \frac{0.270}{0.310} = 0.870,$$

which indicates that X_1 provides a smaller asymptotic variance of the total effect than X_3.

6 CONCLUSION

Instrumental variable method is a powerful tool to estimate total effects when unobserved variables are at presence. In order to identify covariates which satisfy the IV conditions, this paper proposed a method to test whether statistical data is generated by a model involving IVs. We showed that this method is applicable to both continuous case and discrete case. Thus, the results of this paper enable us apply the IV method to wider situations than before.

Finally, we would like to give some further topics. First, as we pointed out in section 1, our method may provide a tighter condition for IVs than Pearl's instrumental inequality. However, we did not provide the detailed discussion because of lack of space. In our opinion, a combination of Pearl's instrumental inequality and our criterion can provide tighter conditions than each of them. This requires further discussion. Second, when two or more covariates satisfy the IV conditions, two important problems occur: one is the IV selection problem, and the other is the robust problem of causal claims through IVs (Pearl, 2004). Although the result of Kuroki and Cai (2004) can be used to solve the former problem, we leave the detailed discussion of applying our results to the latter problem for future research.

ACKNOWLEGDEMENT

The comments of the reviewers on preliminary versions of this paper are acknowledged. This research was supported by the Sumitomo Foundation, No.040705.

REFERENCES

Anderson, T. W. (1986). *Introduction to Multivariate Statistical Analysis*, Second edition. John Wiley & Sons.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects using Instrumental Variables. *Journal of the American Statistical Association*,**91**, 444-472.

Balke, A. and Pearl, J. (1997). Bounds on Treatment Effects from Studies with Imperfect Compliance. *Journal of the American Statistical Association*,**92**, 1171-1176.

Bekker, P. A. and de Leeuw, J. (1987). The Rank of Reduced Dispersion Matrices. *Psychometrika*, **52**, 125-135.

Bollen, K. A. (1989), *Structural Equations with Latent Variables*, John Wiley & Sons.

Bollen, K. A., and Bauer, D. J. (2004). Automating the Selection of Model-Implied Instrumental Variables. *Sociological Methods and Research*, **32**, 425-452.

Bollen, K. A. and Ting, K. (2000). A Tetrad Test for Causal Indicators. *Psychological Methods*, **5**, 3-22.

Bowden, R. J. , and Turkington, D. A. (1984). *Instrumental Variables*, Cambridge University Press.

Brito, C. and Pearl, J. (2002). Generalized Instrumental Variables, *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 85-93.

Chu, T., Scheines, R. and Spirtes, P. (2001). Semi-Instrumental Variables: A Test for Instrument Admissibility *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, 83-90.

Geiger, D., Verma, T. S. and Pearl, J. (1990). Identifying Independence in Bayesian Networks. *Networks*, **20**, 507-534.

Greenland, S. (2000). An Introduction to Instrumental Variables for Epidemiologists, *International Journal of Epidemiology*, **29**,722-729.

Kuroki, M. and Cai, Z. (2004). Selection of Identifiability Criteria for Total Effects by Using Path Diagrams. *Proceeding of the 20th Conference on Uncertainty in Artificial Intelligence*, 333-340.

Kuroki, M., Miyakawa, M. and Cai, Z. (2003). Joint Causal Effect in Linear Structural Equation Model and its Application to Process Analysis. *Proceeding of Ninth International Workshop on Artificial Intelligence and Statistics*, 70-77.

Lauritzen, S. L. (2004). Discussion on Causality, *Scandinavian Journal of Statistics*, **31**, 189-193.

Okuno, T., Katayama, Z., Kamigori, N., Itoh, T., Irikura, N. and Fujiwara, N. (1986). *Kougyou ni okeru Tahenryou Data no Kaiseki* (In Japanese), Nikkagiren, Tokyo.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann.

Pearl, J. (1995). On the Testability of Causal Models with Latent and Instrumental Variables. *Proceeding of the 11th Conference on Uncertainty in Artificial Intelligence*, 435-443.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*, Cambridge University Press.

Pearl, J. (2004). Robustness of Causal Claims, *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 411-420.

Pearl, J. and Tarsi, M. (1986). Structuring Causal Trees, *Journal of Complexity*, **2**, 60-77.

Scheines, R., Cooper, G., Yoo, C. and Chu, T. (2001). Piecewise Linear Instrumental Variable Estimation of Causal Influence, *Proceeding of the 8th International Workshop on Artificial Intelligence and Statistics*, 286-291.

Spearman, C. (1904). General Intelligence Objectively Determined and Measured. *American Journal of Psychology*, **15**, 201-293.

Spearman, C. (1928). Pearson's Contribution to the Theory of Two Factors. *British Journal of Psychology*, **19**, 95-101.

Spirtes, P. , Glymour, C. , and Schienes, R. (1993). *Causation, Prediction, and Search.* Springer-Verlag.

Xu, L. and Pearl, J. (1989). Structuring Causal Tree Models with Continuous Variables, *Uncertainty in Artificial Intelligence*, **3**, 209-219.