Exact Variance Formula of the Estimated Causal Effect on the Variance in Gaussian Linear Structural Equation Models

Taiki Tezuka[a,*], Manabu Kuroki[a]

[a]*Department of Mathematics, Physics, Electrical Engineering and Computer Science, Yokohama National University, 79-1 Tokiwadai, Hodogaya-ku, Yokohama 240-8501 JAPAN.*

Abstract

This paper assumes that cause-effect relationships between random variables can be represented by a Gaussian linear structural equation model and the corresponding directed acyclic graph. Under the situation where we observe a set of random variables that satisfies the back-door criterion, when the ordinary least squares (OLS) method is utilized to estimate the total effect, we formulate the unbiased estimator of the causal effect on the variance (the estimated causal effect on the variance), i.e., the unbiased estimator of the variance of the outcome variable with an external intervention in which a treatment variable is set to a specified constant value. In addition, we provide the variance formula of the estimated causal effect on the variance. The variance formula proposed in this paper is exact, in contrast to those in most previous studies on estimating causal effects.

Keywords: Causal effect, Identification, Path diagram, Structural causal model, Total effect.
2020 MSC: Primary 62D20, Secondary 62H22

1. Introduction

Statistical causal inference using linear structural equation models (linear SEMs) has been widely used to clarify cause-effect relationships between random variables in sociology, economics, biology, etc., and its origin can be traced back to path analysis (Wright, 1923, 1934). Statistical causal inference has been re-developed as the theory of structural causal models (Pearl, 2009).

When a linear SEM is given as a statistical model to describe cause-effect relationships between random variables, the concepts of direct, indirect, and total effects are the important aspects of linear SEMs (Bollen, 1989). According to Bollen (1989), the direct effect is defined as “ those influences unmediated by any other variable in the model," and the indirect effect is defined as “those influences mediated by at least one intervening variable." The total effect is defined as the sum of direct and indirect effects. In the framework of statistical causal inference using linear SEMs, the total effect also means the amount of the change in the expected value of an outcome variable when a treatment variable is changed by one unit by external intervention. To evaluate the total effect, statistical researchers in the field of linear SEMs have provided various identification conditions and estimation methods (e.g., Brito, 2004; Chan and Kuroki, 2010; Chen, 2017; Henckel et al., 2019; Kuroki and Pearl, 2014; Maathuis and Colombo, 2015; Nandy et al., 2017; Pearl, 2009; Perković, 2018; Tian, 2004).

When we wish to characterize the distributional change by the external intervention based on linear SEMs, there is no reason to limit our causal understanding to the change in the expected value of an outcome variable by the external

*Corresponding author. Email address:kuroki-manabu-zm@ynu.ac.jp

intervention. In fact, Hernán and Robins (2022,p.7) stated

> "the average causal effect, defined by a contrast of means of counterfactual outcomes, is the most commonly used population causal effect. However, a population causal effect may also be defined as a contrast of functionals, including medians, variances, hazards, or cdfs of counterfactual outcomes. In general, a population causal effect can be defined as a contrast of any function of the marginal distributions of counterfactual outcomes under different actions or treatment values. For example, the population causal effect on the variance is defined as $var(Y^{a=1}) - var(Y^{a=0})$."

Actually, in practical science, it is important to estimate the change in the expected value of an outcome variable due to the external intervention (the causal effect on the mean). However, it is often necessary to evaluate the variation (variance) of the outcome variable due to the external intervention (the causal effect on the variance) as well. For example, in the field of quality control, in order to suppress a defective rate of products effectively, it is necessary to bring the outcome variable closer to the target value by the external intervention and reduce the variation (or minimize the variance) of the outcome variable as much as possible. In quality control, Kuroki (2008, 2012) and Kuroki and Miyakawa (1999ab) discussed what happens to the variance of the outcome variable when conducting the external intervention. In addition, according to Gische et al.(2021), in medical science, the physician's goal is that the patient's level of blood glucose will be maintained within the euglycemic range (acceptable range) after the treatment (external intervention). Then, the variance of the outcome variable by the external intervention, together with the medical knowledge, plays an important role in constructing the acceptable range to detect a threat to a patient ' s health.

Regarding the estimation accuracy of the causal effect on the variance, when the ordinary least squares (OLS) method is utilized to estimate the total effect, Kuroki and Miyakawa (2003) discussed how the asymptotic variance of the consistent estimator of the causal effect on the variance differs with different sets of random variables that satisfy the back-door criterion (Pearl, 2009). In addition, Shan and Guo (2010) studied the results of Kuroki and Miyakawa (2003) from the perspective of a particular type of external interventions using more than one treatment variable. Shan and Guo (2012) also extended the variable selection criteria provided by Kuroki and Miyakawa (2003) from a deterministic intervention to a stochastic intervention. Kuroki and Nanmo (2020) applied the results of Kuroki and Miyakawa (2003) to predict future values of the outcome variable when conducting the external intervention. Subsequently, Tezuka and Kuroki (2022) pointed out that the existing estimators of the causal effect on the variance (the estimated causal effect on the variance) are the consistent but not unbiased estimators. In addition, they formulated the unbiased estimator of the causal effect on the variance and applied it to the anomaly detection problem. However, they did not provide the exact variance formula of the unbiased estimator of the causal effect on the variance. The estimation accuracy problems are essential issues related to statistical causal inference. This is because the reliable evaluation of the estimation accuracy of the causal effect on the variance plays an important role in the success of statistical data analysis, which aims to evaluate what would happen to the outcome variable when conducting the external intervention based on non-experimental data.

This paper assumes that cause-effect relationships between random variables can be represented by a Gaussian linear SEM and the corresponding directed acyclic graph. Under the situation where we observe a set of random variables that satisfies the back-door criterion, when the OLS method is utilized to estimate the total effect, we formulate the unbiased estimator of the causal effect on the variance, i.e., the unbiased estimator of the variance of the outcome variable with an external intervention in which a treatment variable is set to a specified constant value. In addition, we provide the variance formula of the unbiased estimator of the causal effect on the variance. The variance formula proposed in this paper is exact, in contrast to those in most previous studies on estimating causal effects.

2. Preliminaries

2.1. Gaussian Linear Structural Equation Model

2.2. Graph Terminology

A directed graph is a pair $G = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ is a finite set of vertices. $\boldsymbol{E}$, which is a subset of $\boldsymbol{V} \times \boldsymbol{V}$ of pairs of distinct vertices, is a set of directed edges ($\rightarrow$). If $(a, b) \in E$ for $a, b \in \boldsymbol{V}$, then the G contains the directed edge from vertex a to vertex b (denoted by $a \rightarrow b$). If there is a directed edge from a to b ($a \rightarrow b$), then a is said to be the parent of b and b the child of a. Two vertices are adjacent if there exists a directed edge between them. A path between a and b is a sequence $a = a_0, a_1, \cdots, b = a_m$ of distinct vertices such that a_{i-1} and a_i are adjacent for $i = 1, 2, \cdots, m$. A directed path from a to b is a sequence $a = a_0, a_1, \cdots, b = a_m$ of distinct vertices such that $a_{i-1} \to a_i$ for $i = 1, 2, \cdots, m$. If there exists a directed path from a to b, then a is said to be an ancestor of b and b a descendant of a. When the set of descendants of α is denoted as $de(a)$, the vertices in $V\backslash(de(a) \cup \{a\})$ are said to be the nondescendants of a. If two edges on a path point to a, then a is said to be a collider on the path; otherwise, it is said to be a non-collider on the path. A directed path from a to b, together with the directed edge from b to a, forms a directed cycle. If a directed graph contains no directed cycles, then the graph is said to be a directed acyclic graph (DAG).

2.3. Linear Structural Equation Model

In this paper, it is assumed that cause-effect relationships between random variables can be represented by a Gaussian linear structural equation model (linear SEM) and the corresponding directed acyclic graph (DAG). Then, such a DAG is called a causal path diagram, which is defined as Definition 1. Here, we refer to vertices in the DAG and random variables of the Gaussian linear SEM interchangeably.

Definition 1 (causal path diagram)**.** Consider a DAG $G = (\boldsymbol{V}, \boldsymbol{E})$, for which a set $\boldsymbol{V} = \{V_1, V_2, \cdots, V_m\}$ of random variables and a set $\boldsymbol{E}$ of directed edges are given. Then, the DAG G is called the causal path diagram, if the random variables are generated by a Gaussian linear SEM

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

satisfying the constraints entailed by the DAG G. Here, $\mathrm{pa}(V_i)$ is a set of parents of $V_i \in \boldsymbol{V}$ in the DAG G. In addition, letting $\boldsymbol{0}_m$ be an m-dimensional vector whose i-th element is zero for $i = 1, 2, ..., m$, $\boldsymbol{\epsilon}_v = (\epsilon_{v_1}, \epsilon_{v_2}, \ldots, \epsilon_{v_m})$ denotes a set of random variables, which is assumed to follow the multivariate normal distribution with the mean vector $E(\boldsymbol{\epsilon}_v) = \boldsymbol{0}_m$ and the positive diagonal variance–covariance matrix $\Sigma_{\epsilon_v \epsilon_v}$. In addition, the constant parameters α_{v_i} and $\alpha_{v_i v_j}$ for $i, j = 1, 2, ..., m$ $(i \neq j)$ are referred to as the intercept of V_i and the causal path coefficient (or direct effect) of V_j on V_i, respectively. □

The conditional independence induced by the Gaussian linear SEM (1) can be obtained from the causal path diagram G through the d-separation (Pearl, 2009).

Definition 2 (d-separation)**.** Let $\{X, Y\}$ and $\boldsymbol{Z}$ be the disjoint sets of vertices in the DAG G. If $\boldsymbol{Z}$ blocks every path between distinct vertices X and Y, then $\boldsymbol{Z}$ is said to d-separate X from Y in the DAG G. Here, the path p is said to be blocked by (a possibly empty) set $\boldsymbol{Z}$ if either of the following conditions is satisfied:
(1) p contains at least one non-collider that is in $\boldsymbol{Z}$;
(2) p contains at least one collider that is not in $\boldsymbol{Z}$ and has no descendant in $\boldsymbol{Z}$. □

If $\boldsymbol{Z}$ d-separates X from Y in the causal path diagram G, then X is conditionally independent of Y given $\boldsymbol{Z} = \boldsymbol{z}$ for any value $\boldsymbol{z}$ taken by $\boldsymbol{Z}$ in the corresponding linear SEM (e.g., Pearl, 2009).

2.4. Back-door Criterion

In this paper, for $X, Y \in \boldsymbol{V}$ $(X \neq Y)$, consider the external intervention in which X is set to be the constant value $X = x$ in the Gaussian linear SEM (1), denoted by $\mathrm{do}(X = x)$. According to the framework of the structural causal models (Pearl, 2009), $\mathrm{do}(X = x)$ indicates mathematically that the structural equation for X is replaced by $X = x$ in the Gaussian linear SEM (1).

Let $\boldsymbol{V} = \{X, Y\} \cup \boldsymbol{W}$ be the set of random variables in the causal path diagram G, where $\{X, Y\}$ and $\boldsymbol{W}$ are disjoint. When $f(x, y, \boldsymbol{w})$ and $f(x|\mathrm{pa}(x))$ denote the joint probability distribution of $(X, Y, \boldsymbol{W}) = (x, y, \boldsymbol{w})$ and the conditional probability distribution of $X = x$ given $\mathrm{pa}(X) = \mathrm{pa}(x)$, respectively, the causal effect of X on Y, which is denoted by $f(y|\mathrm{do}(X = x))$, is defined as

$$f(y|\mathrm{do}(X = x)) = \int_w \frac{f(x, y, \boldsymbol{w})}{f(x|\mathrm{pa}(x))} d\boldsymbol{w} \tag{2}$$

(Pearl, 2009). When equation (2) can be uniquely determined from the probability distribution of observed variables, it is said to be identifiable: that is, it can be estimated consistently. Here, in this paper,

$$E(Y|\mathrm{do}(X=x)) = \mu_{y|x} = \int_y y f(y|\mathrm{do}(X=x))\,dy, \quad \mathrm{var}(Y|\mathrm{do}(X=x)) = \sigma_{yy|x} = \int_y (y-\mu_{y|x})^2 f(y|\mathrm{do}(X=x))\,dy \tag{3}$$

are called the causal effect on the mean of Y on $\mathrm{do}(X=x)$ and the causal effect on the variance of Y on $\mathrm{do}(X=x)$, respectively. $E(Y|\mathrm{do}(X=x))$ and $\mathrm{var}(Y|\mathrm{do}(X=x))$ are also called the interventional mean and the interventional variance, respectively, by Gische et al (2021). Then, in the Gaussian linear SEM (1), the first derivative of $E(Y|\mathrm{do}(X=x))$ of Y, namely,

$$\frac{dE(Y|\mathrm{do}(X=x))}{dx} = \tau_{yx} \tag{4}$$

is called the total effect of X on Y. Graphically, the total effect τ_{yx} is interpreted as the total sum of the products of the causal path coefficients on the sequence of directed edges along all directed paths from X to Y. If the total effect τ_{yx} can be uniquely determined from the variance-covariance parameters of observed variables, then it is said to be identifiable; that is, it can be estimated consistently. The interpretation of the total effects in the Gaussian linear SEM (1) via the path analysis (Wright, 1923, 1934) is also discussed by Henckel et al. (2019) and Nandy et al. (2017) in detail.

Let $G_{\underline{X}}$ be the directed graph obtained by deleting all the directed edges emerging from X in the DAG G. Then, the back-door criterion is a well-known identification condition of the causal effect (Pearl, 2009).

Definition 3 (back-door criterion)**.** Let $\{X, Y\}$ and $\boldsymbol{Z}$ be the disjoint subsets of $\boldsymbol{V}$ in the DAG G. If $\boldsymbol{Z}$ satisfies the following conditions relative to an ordered pair (X, Y) in the DAG G, then $\boldsymbol{Z}$ is said to satisfy the back-door criterion relative to (X, Y):

1. no vertex in $\boldsymbol{Z}$ is a descendant of X;
2. $\boldsymbol{Z}$ d-separates X from Y in $G_{\underline{X}}$. □

When $\boldsymbol{Z}$ satisfies the back-door criterion relative to (X, Y) in the causal path diagram G, the causal effect of X on Y is identifiable and is given by

$$f(y|\mathrm{do}(X=x)) = \int_z f(y|x,z)f(z)dz \tag{5}$$

(Pearl, 2009).

Here, we define some notations. For univariates X and Y and a set $\boldsymbol{Z}$ of random variables, let μ_x and μ_y be means of X and Y, respectively. In addition, let σ_{xy}, σ_{xx} and σ_{yy} be the covariance between X and Y, the variance of X and the variance of Y, respectively. When the prime notation ($'$) represents the transpose of a vector or matrix, let Σ_{xz}, Σ_{yz} and Σ_{zz} be the cross covariance vector between X and $\boldsymbol{Z}$ ($\Sigma_{zx} = \Sigma'_{xz}$), the cross covariance vector between Y and $\boldsymbol{Z}$ ($\Sigma_{zy} = \Sigma'_{yz}$) and the variance–covariance matrix of $\boldsymbol{Z}$. Furthermore, for a non-empty set $\boldsymbol{Z}$, let

$$\left.\begin{array}{c} \sigma_{yy.x} = \sigma_{yy} - \dfrac{\sigma_{xy}^2}{\sigma_{xx}}, \ \ \sigma_{xx.z} = \sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}, \ \ \sigma_{xy\cdot z} = \sigma_{xy} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zy}, \ \ \sigma_{yy.z} = \sigma_{yy} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zy} \\ \sigma_{yy.xz} = \sigma_{yy.z} - \dfrac{\sigma_{xy.z}^2}{\sigma_{xx.z}}, \ \ \Sigma_{yz.x} = \Sigma_{yz} - \dfrac{\sigma_{xy}}{\sigma_{xx}}\Sigma_{xz}, \ \ \Sigma_{zy.x} = \Sigma'_{yz.x}, \ \ \Sigma_{zz.x} = \Sigma_{zz} - \dfrac{\Sigma_{zx}\Sigma_{xz}}{\sigma_{xx}} \end{array}\right\}. \tag{6}$$

Then, consider the regression model of Y on X and $\boldsymbol{Z}$

$$Y = \beta_{y.xz} + \beta_{yx.xz}X + B_{yz.xz}\boldsymbol{Z} + \epsilon_{y.xz}, \tag{7}$$

where $\epsilon_{y.xz}$ is a random variable of the regression model (7) that has a normal distribution with zero mean and variance $\sigma_{yy.xz}$, while $\beta_{y.xz}$, $\beta_{yx.xz}$, and $B_{yz.xz}$ are the regression intercept, the regression coefficient of X, and the regression coefficient vector of $\boldsymbol{Z}$ in the regression model (7), respectively. Here, according to the standard assumption of linear regression analysis, in the regression model (7), $\epsilon_{y.xz}$ is assumed to be independent of both X and $\boldsymbol{Z}$. Then, the regression coefficient of X and the regression coefficient vector of $\boldsymbol{Z}$ are given by $\beta_{yx\cdot xz} = \sigma_{xy\cdot z}/\sigma_{xx\cdot z}$ and $B_{yz\cdot xz} = \Sigma_{yz\cdot x}\Sigma_{zz\cdot x}^{-1}$, respectively, when $\sigma_{xx.z} \neq 0$ and $\Sigma_{zz.x}$ is a positive definite matrix.

When a set $\mathbf{Z}$ of observed variables satisfies the back-door criterion relative to (X, Y), then the total effect τ_{yx} is identifiable and is given by $\tau_{yx} = \beta_{yx.xz}$ (Pearl, 2009). Then, according to equation (5), consider the regression model of Y on X and $\mathbf{Z}$, namely, equation (7). Then, $E(Y|\mathrm{do}(X = x))$ and $\mathrm{var}\,(Y|\mathrm{do}(X = x))$ are formulated as

$$E(Y|\mathrm{do}(X = x)) = \mu_{y|x} = \mu_y + \beta_{yx.xz}(x - \mu_x) = \mu_y + \tau_{yx}(x - \mu_x) \tag{8}$$

and

$$\mathrm{var}\,(Y|\mathrm{do}(X = x)) = \sigma_{yy|x} = \sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz}, \tag{9}$$

respectively (Kuroki and Miyakawa, 1999ab, 2003). Here, equation (9) shows that $\mathbf{Z}$ behaves similarly to the random variable by conducting the external intervention $\mathrm{do}(X = x)$, and thus may not reduce the variation of the outcome variable Y by the external intervention (Kuroki, 2012).

To proceed our discussion, we also consider the regression coefficient vector of $\mathbf{Z}$ in the regression model of X on $\mathbf{Z}$

$$X = \beta_{x.z} + B_{xz.z}\mathbf{Z} + \epsilon_{x.z}, \tag{10}$$

where $\epsilon_{x.z}$ is a random variable of the regression model (10) that has a normal distribution with zero mean and variance $\sigma_{xx.z}$, while $\beta_{x.z}$ and $B_{xz.z}$ are the regression intercept and the regression coefficient vector of $\mathbf{Z}$ in the regression model (10), respectively. Here, in the regression model (10), $\epsilon_{x.z}$ is also assumed to be independent of $\mathbf{Z}$. Then, the regression coefficient vector of $\mathbf{Z}$ is denoted by $B_{xz\cdot z} = \Sigma_{xz}\Sigma_{zz}^{-1}$ when Σ_{zz} is a positive definite matrix.

3. Results

Let $\hat{\mu}_x$ and $\hat{\mu}_y$ be the sample means of X and Y, respectively. In addition, let s_{xx}, s_{yy}, s_{xy}, S_{zz}, S_{xz} and S_{yz} be the sum-of-squares of X, the sum-of-squares of Y, the sum-of cross-products between X and Y, the sum-of-squares matrix of $\mathbf{Z}$, the sum-of-cross-products vector between X and $\mathbf{Z}$ ($S_{zx} = S'_{xz}$), and the sum-of-cross-products vector between Y and $\mathbf{Z}$ ($S_{zy} = S'_{yz}$), respectively. Based on the notation, let

$$\left.\begin{array}{c} s_{yy.x} = s_{yy} - \dfrac{s_{xy}^2}{s_{xx}}, \quad s_{xx.z} = s_{xx} - S_{xz}S_{zz}^{-1}S_{zx}, \quad s_{xy.z} = s_{xy} - S_{xz}S_{zz}^{-1}S_{zy}, \quad s_{yy.z} = s_{yy} - S_{yz}S_{zz}^{-1}S_{zy}, \\ s_{yy.xz} = s_{yy.z} - \dfrac{s_{xy.z}^2}{s_{xx.z}}, \quad S_{yz.x} = S_{yz} - \dfrac{s_{xy}}{s_{xx}}S_{xz}, \quad S_{zy.x} = S'_{yz.x}, \quad S_{zz.x} = S_{zz} - \dfrac{S_{zx}S_{xz}}{s_{xx}} \end{array}\right\}. \tag{11}$$

Then, through the ordinary least squares (OLS) method, the unbiased estimators of $\beta_{yx.xz}$, $B_{xz.z}$ and $B_{yz.xz}$ of equations (7) and (10) are given by $\hat{\beta}_{yx.xz} = s_{xy.z}/s_{xx.z}$, $\hat{B}_{xz.z} = S_{xz}S_{zz}^{-1}$ and $\hat{B}_{yz.xz} = S_{yz.x}S_{zz.x}^{-1}$, respectively, when S_{zz} and $S_{zz.x}^{-1}$ are positive definite matrices. Here, letting n and q be the sample size and the number of random variables in $\mathbf{Z}$, respectively, for $q < n - 2$,

$$\hat{\sigma}_{yy.xz} = \frac{s_{yy.xz}}{n - q - 2}, \quad \hat{\Sigma}_{zz} = \frac{1}{n-1}S_{zz} \tag{12}$$

are also unbiased estimators of $\sigma_{yy.xz}$ and Σ_{zz}, respectively.

Under the random sampling, when the total effect τ_{yx} is estimated as $\hat{\tau}_{yx} = \hat{\beta}_{yx\cdot xz}$ through the OLS method in the regression model (7), the exact variance of $\hat{\beta}_{yx\cdot xz}$ is given by

$$\mathrm{var}\left(\hat{\beta}_{yx\cdot xz}\right) = \frac{1}{n - q - 3}\frac{\sigma_{yy\cdot xz}}{\sigma_{xx\cdot z}} \tag{13}$$

for $q < n - 3$ (e.g., Kuroki and Cai, 2004).

The following theorem holds:

Theorem 1. *Under the Gaussian linear SEM (1), suppose that* $\mathbf{Z}$ *satisfies the back-door criterion relative to* (X, Y) *in the causal path diagram G. When the OLS method is utilized to evaluate the statistical parameters in equations (8) and (9), the unbiased estimators of the causal effect on the mean* $\mu_{y|x} = E(Y|do(X = x))$ *and the causal effect on the variance* $\sigma_{yy|x} = var\,(Y|do(X = x))$ *are given by*

$$\hat{\mu}_{y|x} = \hat{\mu}_y + \hat{\beta}_{yx.xz}(x - \hat{\mu}_x) \tag{14}$$

$$\hat{\sigma}_{yy|x} = \hat{\sigma}_{yy.xz}\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right) + \hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}, \tag{15}$$

respectively. $\hat{\mu}_{y|x}$ *and* $\hat{\sigma}_{yy|x}$ *are called the estimated causal effect on the mean of Y on do(X = x) and estimated causal effect on the variance of Y on do(X = x), respectively. In addition, for* $q < n-5$*, the variances* $var(\hat{\mu}_{y|x})$ *of* $\hat{\mu}_{y|x}$ *and* $var(\hat{\sigma}_{yy|x})$ *of* $\hat{\sigma}_{yy|x}$ *are given by*

$$var\left(\hat{\mu}_{y|x}\right) = \frac{1}{n}\left(\sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz}\right) + \frac{\sigma_{yy.xz}}{(n-q-3)\sigma_{xx.z}}\left((x-\mu_x)^2 + \frac{\sigma_{xx}}{n}\right), \tag{16}$$

$$\begin{aligned}var(\hat{\sigma}_{yy|x}) = {}& \frac{2(B_{yz.xz}\Sigma_{zz}B'_{yz.xz})^2}{n-1} + \frac{2\sigma^2{}_{yy.xz}}{n-q-2}\left(\left(1-\frac{q}{n-1}\right)^2 - 2\left(1-\frac{q}{n-1}\right)\frac{q\sigma_{xx.z}+(n-1)B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-1)(n-q-3)\sigma_{xx.z}}\right.\\ &\left.+E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{(n-1)s_{xx.z}}\right)^2\right)\right) + \frac{2\sigma^2_{yy.xz}}{(n-1)^2}\left(q + 2\frac{q\sigma_{xx.z}+(n-1)B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)\sigma_{xx.z}} + E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right)\right)\\ &+\frac{4\sigma_{yy.xz}}{(n-1)^2}\left((n-1)B_{yz.xz}\Sigma_{zz}B'_{yz.xz} + E\left(\frac{(B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2}{s_{xx.z}}\right)\right),\end{aligned} \tag{17}$$

respectively, where

$$\begin{aligned}E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right) = {}& \frac{2q\sigma^2_{xx.z}+4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B_{xz.z}}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} + \frac{2(n-1)(B_{xz.z}\Sigma_{zz}B_{xz.z})^2}{(n-q-3)(n-q-5)\sigma^2_{xx.z}}\\ &+\frac{(q\sigma_{xx.z}+(n-1)B_{xz.z}\Sigma_{zz}B_{xz.z})^2}{(n-q-3)(n-q-5)\sigma^2_{xx.z}}\end{aligned} \tag{18}$$

$$E\left(\frac{(B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2}{s_{xx.z}}\right) = \frac{(n-1)(n(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2 + (B_{xz.z}\Sigma_{zz}B'_{xz.z})(B_{yz.xz}B'_{xz.z})^2 + \sigma_{xx.z}B_{yz.xz}\Sigma_{zz}B'_{yz.xz})}{(n-q-3)\sigma_{xx.z}}. \tag{19}$$

□

Equations (14) and (15) are given by Kuroki and Nanmo (2020) and Tezuka and Kuroki (2022), respectively. Equation (16) is also given by Kuroki and Nanmo (2020). The derivation of equation (17), which is one of the new results, is provided in Appendix. Here, we also provide the derivation of equation (15) in the Appendix because Tezuka and Kuroki (2022) is written in Japanese.

For the large sample size n such as $n^{-2} \simeq 0$, the consistent estimator $\hat{\sigma}_{yy|x}$ of $\sigma_{yy|x}$ can be given by

$$\hat{\sigma}_{yy|x} = \hat{\sigma}_{yy.xz} + \hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}, \tag{20}$$

which shows that equation (20) is larger than equation (15). In addition, the asymptotic variance of $\hat{\sigma}_{yy|x}$, a.var($\hat{\sigma}_{yy|x}$), is given by

$$\begin{aligned}\text{a.var}(\hat{\sigma}_{yy|x}) &= \frac{2\sigma^2_{yy.xz}}{n} + \frac{2(B_{yz.xz}\Sigma_{zz}B_{yz.xz})^2}{n} + \frac{4\sigma_{yy.xz}}{n}\left(B_{yz.xz}\Sigma_{zz}B'_{yz.xz} + \frac{(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2}{\sigma_{xx.z}}\right)\\ &= \frac{2}{n}\left(\sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz}\right)^2 + \frac{4\sigma_{yy.xz}}{n\sigma_{xx.z}}(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2.\end{aligned} \tag{21}$$

Here, when we let $\beta_{yx.x} = \sigma_{xy}/\sigma_{xx}$, $\beta_{yx.xz} = \tau_{yx}$, $B_{xz.z} = \Sigma_{xz}\Sigma^{-1}_{zz}$ and the covariance between X and equation (7) leads to

$$\sigma_{xy} = \beta_{yx.xz}\sigma_{xx} + B_{yz.xz}\Sigma_{zx} = \tau_{yx}\sigma_{xx} + B_{yz.xz}\Sigma_{zx}, \tag{22}$$

which provides

$$B_{yz.xz}\Sigma_{zz}B'_{xz.z} = B_{yz.xz}\Sigma_{zx} = (\beta_{yx.x} - \tau_{yx})\sigma_{xx} \tag{23}$$

and

$$\sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz} = \sigma_{yy.x} - B_{yz.xz}\Sigma_{zz.x}B'_{yz.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz} = \sigma_{yy.x} + \frac{(B_{yz.xz}\Sigma_{zx})^2}{\sigma_{xx}} = \sigma_{yy.x} + (\beta_{yx.x} - \tau_{yx})^2\sigma_{xx}. \quad (24)$$

From equation (24), the first term of equation (21), which is equivalent to equation (9), does not depend on the selection of the set $\mathbf{Z}$ of random variables that satisfies the back-door criterion (Kuroki, 2008, 2012). In addition, $B_{yz.xz}\Sigma_{zz}B'_{xz.z}$ in the second term of equation (21) does not depend on the selection of the set $\mathbf{Z}$ of random variables. Thus, the difference between selected sets of random variables depends on $\sigma_{yy.xz}/\sigma_{xx.z}$ in the second term of equation (21). From this consideration, letting $\hat{\sigma}_{yy|x,z}$ be the estimated causal effect on the variance of Y on do($X = x$) to emphasize that $\mathbf{Z}$ is utilized to estimate equation (9), the following theorem is the extension of the variable selection criterion given by Kuroki and Miyakawa (2003), from the univariate case to the multivariate case.

Theorem 2. *Under the Gaussian linear SEM (1), suppose that sets $\mathbf{Z}_1$ and $\mathbf{Z}_2$ of random variables satisfy the back-door criterion relative to (X, Y) in the causal path diagram G. When the OLS method is utilized to evaluate the statistical parameters in equations (8) and (9), if $\mathbf{Z}_2$ d-separates X from $\mathbf{Z}_1$, then*

$$a.var(\hat{\sigma}_{yy|x.z_1,z_2}) \leq a.var(\hat{\sigma}_{yy|x.z_2}) \quad (25)$$

holds, and if $\{X\} \cup \mathbf{Z}_1$ d-separates Y from $\mathbf{Z}_2$, then

$$a.var(\hat{\sigma}_{yy|x.z_1}) \leq a.var(\hat{\sigma}_{yy|x.z_1,z_2}) \quad (26)$$

holds. □

The proof of Theorem 2 is trivial from the following lemma given by Kuroki and Cai (2004):

Lemma 1. *When $\{X, Y\} \cup \mathbf{Z}_1 \cup \mathbf{Z}_2$ follows a multivariate normal distribution, if X is conditionally independent of $\mathbf{Z}_1$ given $\mathbf{Z}_2$, then*

$$\frac{\sigma_{yy\cdot xz_1z_2}}{\sigma_{xx\cdot z_1z_2}} \leq \frac{\sigma_{yy\cdot xz_2}}{\sigma_{xx\cdot z_2}} \quad (27)$$

holds, and if Y is conditionally independent of $\mathbf{Z}_2$ given $\{X\} \cup \mathbf{Z}_1$, then

$$\frac{\sigma_{yy\cdot xz_1}}{\sigma_{xx\cdot z_2}} \leq \frac{\sigma_{yy\cdot xz_1z_2}}{\sigma_{xx\cdot z_1z_2}} \quad (28)$$

holds.

4. Numerical Experiments

This section will report numerical experiments conducted to examine statistical properties of the estimated causal effect on the variance for sample sizes $n = 10, 25, 50$, and 100. For simplicity, consider the DAG depicted in Figure 1 and the Gaussian linear SEM in the form of

$$Y = \alpha_{yx}X + \alpha_{yz_1}Z_1 + \epsilon_y, \quad X = \alpha_{xz_2}Z_2 + \epsilon_x, \quad Z_1 = \alpha_{z_1z_2}Z_2 + \epsilon_{z_1}, \quad Z_2 = \epsilon_{z_2}, \quad (29)$$

where ϵ_x, ϵ_y, ϵ_{z_1} and ϵ_{z_2} independently follow a normal distribution with mean zero. The matrices of the path coefficients of X, Y, Z_1, and Z_2 shown in Table 1 are utilized for our purpose. In this situation, $\mathbf{Z} = \{Z_1\}, \{Z_2\}$ and $\{Z_1, Z_2\}$ satisfy the back-door criterion relative to (X, Y). Cases 1 and 2 represent situations where the empty set satisfies the back-door criterion relative to (X, Y). Because X is independent of $\{Z_1, Z_2\}$ in Case 1, we obtain $\tau_{yx} = \beta_{yx.x} = \beta_{yx.xz}$ for $\mathbf{Z}$, and the information about $\mathbf{Z}$ would asymptotically improve the estimation accuracy of the total effect τ_{yx} (Kuroki and Cai, 2004). In Case 2, because Y is conditionally independent of $\mathbf{Z}$ given X, we also obtain $\tau_{yx} = \beta_{yx.x} = \beta_{yx.xz}$. However, the information about $\mathbf{Z}$ does not asymptotically improve the estimation accuracy of the total effect τ_{yx} (Kuroki and Cai, 2004). Cases 3 and 4 represent situations in which $\mathbf{Z}$ satisfies the back-door criterion relative to (X, Y); however, parametric cancellation occurs (Cox and Wermuth, 2014), where $\beta_{yx.x} = 0$ and $\tau_{yx} = \beta_{yx.xz} \neq 0$ hold

Table 1. Path Coefficients

	Case1				Case2				Case3			
	Y	X	Z_1	Z_2	Y	X	Z_1	Z_2	Y	X	Z_1	Z_2
Y	-	0.7000	0.7000	0.0000	-	0.7000	0.0000	0.0000	-	-0.3430	0.7000	0.0000
X	-	-	0.0000	0.0000	-	-	0.0000	0.7000	-	-	0.0000	0.7000
Z_1	-	-	-	0.7000	-	-	-	0.7000	-	-	-	0.7000

	Case4				Case5			
	Y	X	Z_1	Z_2	Y	X	Z_1	Z_2
Y	-	0.0000	0.7000	0.0000	-	-1.9697	2.5303	0.0000
X	-	-	0.0000	0.7000	-	-	0.0000	0.9900
Z_1	-	-	-	0.7000	-	-	-	0.9900

and (21) when $\{Z_1, Z_2\}$ is selected is larger than equations (17) and (21) when Z_1 is selected, respectively. This implies that the relationships are consistent with the results obtained by Theorem 2. In contrast, in Case 2 with the sample size $n \leq 50$, equation (17) when $\{Z_1, Z_2\}$ is selected is larger than equation (17) when Z_2 is selected, which shows that the relationships are different from the results obtained by Theorem 2. Thus, it seems that the difference between the estimation accuracy by the selected variables depends on not only the sample size but also the multicollinearity between X and $\mathbf{Z}$: Theorem 2 holds for large sample sizes when X is highly correlated with $\mathbf{Z}$.

Third, comparing the empirical variances with the variance formula, equation (17) is relatively close to the empirical variances of the unbiased estimator for any sample size. In contrast, when X is correlated with $\mathbf{Z}$, the asymptotic variance (21) is not close to the empirical variances of the consistent estimator for the small sample sizes in each case. Especially, the differences between the asymptotic variance (21) and the empirical variances of the consistent estimator are significant when X is correlated with $\mathbf{Z}$. However, they become close as the sample size is larger.

Finally, for each case, it seems that both unbiased and consistent estimators are highly skewed and heavy-tailed in the small sample size, but converge to the normal distributions slowly as the sample sizes are larger. Especially, when X is correlated with $\mathbf{Z}$, both unbiased and consistent estimators take large positive/negative values in the small sample size, which implies that these estimators are unstable under the multicollinearity with the small sample size.

5. Conclusion

In this paper, when causal knowledge is available in the form of a Gaussian linear SEM with the corresponding DAG, we considered a situation where the causal effect can be estimated based on the back-door criterion. Under this situation, we formulated the unbiased estimator of the causal effect on the variance with the exact variance. The estimated causal effect on the variance proposed by Kuroki and Miyakawa (2003) and Kuroki and Nanmo (2020) is consistent but not unbiased. Under the small sample size, the use of the consistent estimator may lead to misleading findings in statistical causal inference. The proposed estimator would help us avoid the problem, and the results of this paper would help statistical practitioners to predict appropriately what would happen to the outcome variable when conducting the external intervention.

Future work should involve extending our results to (i) a joint intervention that combines several single interventions and (ii) an adaptive control in which the treatment variable is assigned a value based on some covariates. In addition, the numerical experiments show that the proposed unbiased estimator has the drawback that it can take a negative value in the small sample size. One of our suggestion to solve the problem is to use the $\max\{0, \hat{\sigma}_{yy|x}\}$ but not $\hat{\sigma}_{yy|x}$ to evaluate the causal effect on the variance. However, noting that $\max\{0, \hat{\sigma}_{yy|x}\}$ is not an unbiased estimator, it would also be future work to develop the more efficient estimator of the causal effect on the variance.

Acknowledgement

This research was partially supported by Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science, Grant Number 19K11856.

Appendix: Proof of Theorem 1

Letting D_x and D_z denote the datasets of X and $\mathbf{Z}$, respectively, from the law of total variance (Weiss et al, 2006, pp.385-386), given $D_x \cup D_z$, we have

$$\mathrm{var}(\hat{\sigma}_{yy|x}) = \mathrm{var}(E(\hat{\sigma}_{yy|x}|D_x, D_z)) + E(\mathrm{var}(\hat{\sigma}_{yy|x}|D_x, D_z)), \tag{30}$$

where $E(\cdot|D_x, D_z)$ and $\mathrm{var}(\cdot|D_x, D_z)$ indicates that conditional expectation and variance given $D_x \cup D_z$, respectively. Then, in order to derive the explicit expression of the exact variance formula of the estimated causal effect on the variance $\hat{\sigma}_{yy|x}$ of Y on $\mathrm{do}(X = x)$, we calculate the first term $\mathrm{var}(E(\hat{\sigma}_{yy|x}|D_x, D_z))$ and the second term $E(\mathrm{var}(\hat{\sigma}_{yy|x}|D_x, D_z))$ of equation (30) separately.

Step 1:Derivation of var$(E(\hat{\sigma}_{yy|x}|D_x, D_z))$

Regarding the second term of the right hand side of equation (15), note that we derive

$$
\begin{aligned}
E(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z) = E(\mathrm{tr}(\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}\hat{B}_{yz.xz})|D_x, D_z) &= \mathrm{tr}\left(\hat{\Sigma}_{zz}(\sigma_{yy.xz}S^{-1}_{zz.x} + B'_{yz.xz}B_{yz.xz})\right) \\
&= \sigma_{yy.xz}\mathrm{tr}(\hat{\Sigma}_{zz}S^{-1}_{zz.x}) + B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz} \qquad (31)
\end{aligned}
$$

by Mathai and Provost (1992, p.53) and the basic formula of the variance-covariance matrix

$$
\begin{aligned}
\mathrm{var}(\hat{B}_{yz.xz}|D_x, D_z) &= E(\hat{B}'_{yz.xz}\hat{B}_{yz.xz}|D_x, D_z) - E(\hat{B}'_{yz.xz}|D_x, D_z)E(\hat{B}_{yz.xz}|D_x, D_z) = E(\hat{B}'_{yz.xz}\hat{B}_{yz.xz}|D_x, D_z) - B'_{yz.xz}B_{yz.xz} \\
&= \sigma_{yy.xz}S^{-1}_{zz.x}, \qquad (32)
\end{aligned}
$$

where tr(A), which is the trace of a square matrix A, represents as the total sum of elements on the main diagonal of the square matrix A. Thus, noting that equation (11), $\hat{\sigma}_{yy.xz}$, is the unbiased estimator of $\sigma_{yy.xz}$, we have

$$
\begin{aligned}
E(\hat{\sigma}_{yy|x}|D_x, D_z) &= E\left(\hat{\sigma}_{yy.xz}\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right)|D_x, D_z\right) + E(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z) \\
&= \sigma_{yy.xz}\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right) + \mathrm{tr}(\hat{\Sigma}_{zz}S^{-1}_{zz.x})\right) + B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}. \qquad (33)
\end{aligned}
$$

Here, from Sherman–Morrison formula (Sherman and Morrison, 1950), $S^{-1}_{zz.x}$ can be expressed as

$$
S^{-1}_{zz.x} = \left(S_{zz} - \frac{S_{zx}S_{xz}}{s_{xx}}\right)^{-1} = S^{-1}_{zz} + \frac{S^{-1}_{zz}S_{zx}S_{xz}S^{-1}_{zz}}{s_{xx.z}}. \qquad (34)
$$

Thus, from equation (12), noting that $\hat{\Sigma}_{zz}$ is the unbiased estimator of Σ_{zz}, we derive

$$
\begin{aligned}
\mathrm{tr}(\hat{\Sigma}_{zz}S^{-1}_{zz.x}) &= \frac{1}{n-1}\mathrm{tr}(S_{zz}S^{-1}_{zz.x}) = \frac{1}{n-1}\mathrm{tr}\left(I_{q,q} + \frac{S_{zx}S_{xz}S^{-1}_{zz}}{s_{xx.z}}\right) = \frac{1}{n-1}\left(q + \frac{S_{xz}S^{-1}_{zz}S_{zx}}{s_{xx.z}}\right) \\
&= \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right), \qquad (35)
\end{aligned}
$$

where $I_{q,q}$ is the $q \times q$ identity matrix. Thus, since we have

$$
E(\hat{\sigma}_{yy|x}|D_x, D_z) = \sigma_{yy.xz} + B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz} \qquad (36)
$$

from equation (31) together with equation (35), we derive

$$
E(\hat{\sigma}_{yy|x}) = E(E(\hat{\sigma}_{yy|x}|D_x, D_z)) = \sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz} \qquad (37)
$$

and

$$
\mathrm{var}(E(\hat{\sigma}_{yy|x}|D_x, D_z)) = \mathrm{var}(B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}). \qquad (38)
$$

Equation (37) shows that $\hat{\sigma}_{yy|x}$ is the unbiased estimator of the causal effect on the variance of Y on do$(X = x)$ (Tezuka and Kuroki, 2022).

Here, noting that $(n-1)\hat{\Sigma}_{zz}$ follows the Wishart distribution with the $n-1$ degrees of freedom and parameter Σ_{zz} and

$$
\frac{(n-1)B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}}{B_{yz.xz}\Sigma_{zz}B'_{yz.xz}} \qquad (39)
$$

follows the chi-squared distribution with $n-1$ degrees of freedom (Seber, 2008, p.466), the variance is given by

$$
\mathrm{var}\left(\frac{(n-1)B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}}{B_{yz.xz}\Sigma_{zz}B'_{yz.xz}}\right) = 2(n-1), \qquad (40)
$$

i.e., we have

$$
\mathrm{var}(E(\hat{\sigma}_{yy|x}|D_x, D_z)) = \mathrm{var}\left(B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}\right) = \frac{2(B_{yz.xz}\Sigma_{zz}B'_{yz.xz})^2}{n-1}. \qquad (41)
$$

Step 2: Derivation of $E(\mathbf{var}(\hat{\sigma}_{yy|x}|D_x, D_z))$

Noting that $\hat{\sigma}_{yy.xz}$ and $(\hat{\beta}_{yx.xz}, \hat{B}'_{yz.xz})'$ are independent of each other given D_x and D_z (e.g., Mardia et al, 1979), since

$$\frac{(n-q-2)\hat{\sigma}_{yy.xz}}{\sigma_{yy.xz}} \tag{42}$$

follows the chi-squared distribution with $n-q-2$ degrees of freedom, we have

$$\begin{aligned}
\mathrm{var}(\hat{\sigma}_{yy|x}|D_x, D_z) &= \mathrm{var}(\hat{\sigma}_{yy.xz}|D_x, D_z)\left(1-\frac{1}{n-1}\left(q+\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right)^2 + \mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z)\\
&= \frac{2\sigma^2{}_{yy.xz}}{n-q-2}\left(1-\frac{1}{n-1}\left(q+\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right)^2 + \mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z)\\
&= \frac{2\sigma^2{}_{yy.xz}}{n-q-2}\left(\left(1-\frac{q}{n-1}\right)^2 - 2\left(1-\frac{q}{n-1}\right)\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{(n-1)s_{xx.z}} + \left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{(n-1)s_{xx.z}}\right)^2\right) + \mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z).
\end{aligned} \tag{43}$$

Step 2-1: Derivation of $E\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)$

Regarding the first term of equation (43), since $\hat{B}_{xz.z}$ and $s_{xx.z}$ are independent of each other given D_z (e.g., Mardia et al, 1979), noting that $s_{xx.z}/\sigma_{xx.z}$ follows the chi-squared distribution with $n-q-1$ degrees of freedom, we have

$$E\left(\frac{1}{s_{xx.z}}|D_z\right) = \frac{1}{(n-q-3)\sigma_{xx.z}}. \tag{44}$$

Thus, we have

$$\begin{aligned}
E\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right) &= E\left(E\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}|D_z\right)\right) = E\left(E(\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}|D_z)E\left(\frac{1}{s_{xx.z}}|D_z\right)\right)\\
&= \frac{\sigma_{xx.z}E(\mathrm{tr}(S_{zz}S_{zz}^{-1})) + B_{xz.z}E(S_{zz})B'_{xz.z}}{(n-q-3)\sigma_{xx.z}} = \frac{q\sigma_{xx.z}+(n-1)B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)\sigma_{xx.z}}
\end{aligned} \tag{45}$$

from

$$\mathrm{var}(\hat{B}_{xz.z}) = E(\hat{B}'_{xz.z}\hat{B}_{xz.z}) - B'_{xz.z}B_{xz.z} = \sigma_{xx.z}S_{zz}^{-1}. \tag{46}$$

Step 2-2: Derivation of $E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right)$

Similar to Step 2-1, from

$$E\left(\frac{1}{s_{xx.z}^2}\right) = \frac{1}{(n-q-3)(n-q-5)\sigma_{xx.z}^2}, \tag{47}$$

since $\hat{B}_{xz.z}$ and $\hat{\sigma}_{xx.z}$ are independent of each other given D_z (e.g., Mardia et al, 1979), we derive

$$\begin{aligned}
E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right) &= E\left(E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2|D_z\right)\right) = E\left(E((\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z})^2|D_z)E\left(\frac{1}{s_{xx.z}^2}|D_z\right)\right)\\
&= \frac{E\left(E((\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z})^2|D_z)\right)}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} = \frac{E\left(\mathrm{var}\left(\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}|D_z\right)\right)}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} + \frac{E(E(\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}|D_z)^2)}{(n-q-3)(n-q-5)\sigma_{xx.z}^2}.
\end{aligned} \tag{48}$$

From Seber (2008, p.438) and equation (12), $E\left(\mathrm{var}\left(\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}|D_z\right)\right)$ is given by

$$\begin{aligned}
E\left(\mathrm{var}\left(\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}|D_z\right)\right) &= 2\sigma_{xx.z}^2E(\mathrm{tr}(S_{zz}S_{zz}^{-1}S_{zz}S_{zz}^{-1})) + 4E(\sigma_{xx.z}B_{xz.z}S_{zz}S_{zz}^{-1}S_{zz}B'_{xz.z})\\
&= 2q\sigma_{xx.z}^2 + 4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}.
\end{aligned} \tag{49}$$

Again, from

$$\mathrm{var}(B_{xz.z}S_{zz}B'_{xz.z}) = 2(n-1)(B_{xz.z}\Sigma_{zz}B'_{xz.z})^2 \tag{50}$$

by Seber (2008, p.466), we have

$$\begin{aligned}
E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right) &= E\left(E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2 |D_z\right)\right)\\
&= \frac{2q\sigma^2_{xx.z}+4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} + \frac{E((q\sigma_{xx.z}+B_{xz.z}S_{zz}B'_{xz.z})^2)}{(n-q-3)(n-q-5)\sigma^2_{xx.z}}\\
&= \frac{2q\sigma^2_{xx.z}+4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} + \frac{\mathrm{var}(B_{xz.z}S_{zz}B'_{xz.z})}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} + \frac{E(q\sigma_{xx.z}+B_{xz.z}S_{zz}B'_{xz.z})^2}{(n-q-3)(n-q-5)\sigma^2_{xx.z}}\\
&= \frac{2q\sigma^2_{xx.z}+4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} + \frac{2(n-1)(B_{xz.z}\Sigma_{zz}B'_{xz.z})^2}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} + \frac{(q\sigma_{xx.z}+(n-1)B_{xz.z}\Sigma_{zz}B'_{xz.z})^2}{(n-q-3)(n-q-5)\sigma^2_{xx.z}}.
\end{aligned} \tag{51}$$

Step 2-3: Derivation of var$(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z)$

Regarding the second term of equation (43), from Mathai and Provost (1992, p.53), we have

$$\mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z) = \frac{2\sigma^2_{yy.xz}}{(n-1)^2}\mathrm{tr}(S_{zz}S^{-1}_{zz.x}S_{zz}S^{-1}_{zz.x}) + \frac{4\sigma_{yy.xz}}{(n-1)^2}B_{yz.xz}S_{zz}S^{-1}_{zz.x}S_{zz}B'_{yz.xz}. \tag{52}$$

From equation (29) and $\hat{B}_{xz.z} = S_{xz}S^{-1}_{zz}$, we have

$$\begin{aligned}
E(\mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z)) &= E\left(\frac{2\sigma^2_{yy.xz}}{(n-1)^2}\mathrm{tr}\left(\left(S_{zz}+\frac{S_{zx}S_{xz}}{s_{xx.z}}\right)S^{-1}_{zz.x}\right)\right) + \frac{4\sigma_{yy.xz}}{(n-1)^2}B_{yz.xz}E\left(S_{zz}+\frac{S_{zx}S_{xz}}{s_{xx.z}}\right)B'_{yz.xz}\\
&= \frac{2\sigma^2_{yy.xz}}{(n-1)^2}\left(q+2E\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)+E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right)\right)\\
&\quad +\frac{4\sigma_{yy.xz}}{(n-1)^2}\left((n-1)B_{yz.xz}\Sigma_{zz}B'_{yz.xz}+E\left(\frac{(B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2}{s_{xx.z}}\right)\right)
\end{aligned} \tag{53}$$

Here, from the law of total variance (Weiss et al, 2006, pp.385-386), we have

$$\begin{aligned}
E\left(E\left((B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2|D_z\right)\right) &= E\left(\mathrm{var}\left(B_{yz.xz}S_{zz}\hat{B}'_{xz.z}|D_z\right)+E\left((B_{yz.xz}S_{zz}\hat{B}'_{xz.z})|D_z\right)^2\right)\\
&= \sigma_{xx.z}B_{yz.xz}E(S_{zz})B'_{yz.xz}+E((B_{yz.xz}S_{zz}B'_{xz.z})^2) = (n-1)\sigma_{xx.z}B_{yz.xz}\Sigma_{zz}B'_{yz.xz}+E((B_{yz.xz}S_{zz}B'_{xz.z})^2)
\end{aligned} \tag{54}$$

Thus, from equations (21) and (21). Finally, from Seber (2008,p.467), we have

$$\begin{aligned}
E((B_{yz.xz}S_{zz}B'_{xz.z})^2) &= B_{yz.xz}E(S_{zz}B'_{xz.z}B_{xz.z}S_{zz})B'_{yz.xz}\\
&= ((n-1)+(n-1)^2)(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2+(n-1)(B_{xz.z}\Sigma_{zz}B'_{xz.z})(B_{yz.xz}B'_{xz.z})^2.
\end{aligned} \tag{55}$$

Based on the above derivation, we derive the exact variance formula of the estimated causal effect on the variance of Y on do$(X = x)$.

References

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

[2] Brito, C. (2004). Graphical methods for identification in structural equation models. Ph.D. Thesis, Department of Computer Science Department, UCLA.

[3] Chan, H. and Kuroki, M. (2010). Using descendants as instrumental variables for the identification of direct causal effects in linear SEMs. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 73-80.

[4] Chen, B. R. (2017). Graphical methods for linear structural equation modeling. Ph.D. Thesis, Department of Computer Science Department, UCLA.
[5] Cox, D. R. and Wermuth, N. (2014). *Multivariate Dependencies: Models, Analysis and Interpretation*, Chapman and Hall/CRC.
[6] Gische, C., West, S. G. and Voelkle, M. C. (2021). Forecasting causal effects of interventions versus predicting future outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, **28**, 475-492.
[7] Henckel, L., Perković, E. and Maathuis, M. H. (2019). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. arXiv preprint arXiv:1907.02435.
[8] Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*, Chapman & Hall/CRC.
[9] Kuroki, M. (2008). The evaluation of causal effects on the variance and its application to process analysis. *Journal of the Japanese Society for Quality Control*, **38**, 373-384.
[10] Kuroki, M. (2012). Optimizing an external intervention using a structural equation model with an application to statistical process analysis. *Journal of Applied Statistics*, **39**, 673-694.
[11] Kuroki, M. and Cai, Z. (2004). Selection of identifiability criteria for total effects by using path diagrams. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 333-340.
[12] Kuroki, M. and Miyakawa, M. (1999a). Estimation of causal effects in causal diagrams and its application to process analysis. *Journal of the Japanese Society for Quality Control*, **29**, 70-80.
[13] Kuroki, M. and Miyakawa, M. (1999b). Estimation of conditional intervention effect in adaptive control. *Journal of the Japanese Society for Quality Control*, **29**, 237-245.
[14] Kuroki, M. and Miyakawa, M. (2003). Covariate selection for estimating the causal effect of external interventions using causal diagrams. *Journal of the Royal Statistical Society, Series B*, **65**, 209-222.
[15] Kuroki, M. and Nanmo, H. (2020). Variance formulas for estimated causal effect on the mean and predicted outcome with external intervention based on the back-door criterion in linear structural equation models. *AStA Advances in Statistical Analysis*, **104**, 667-685.
[16] Kuroki, M. and Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika*, **101**, 423-437.
[17] Maathuis, M. H. and Colombo, D. (2015). A generalized back-door criterion. *Annals of Statistics*, **43**, 1060-1088.
[18] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press.
[19] Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables: Theory and Applications*, CRC Press.
[20] Nandy, P., Maathuis, M. H. and Richardson, T. S. (2017) . Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *Annals of Statistics*, **45**, 647–674.
[21] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference, the 2nd Edition*, Cambridge University Press.
[22] Perković, E. (2018). Graphical characterizations of adjustment sets. Ph.D. Thesis, Department of Mathematics, ETH Zurich.
[23] Seber, G. A. (2008). *A matrix handbook for statisticians*, John Wiley & Sons.
[24] Shan, N. and Guo, J. (2010). Covariate selection for identifying the effects of a particular type of conditional plan using causal networks. *Frontiers of Mathematics in China*, **5**, 687-700.
[25] Shan, N. and Guo, J. (2012). Covariate selection for identifying the causal effects of stochastic interventions using causal networks. *Journal of Statistical Planning and Inference*, **142**, 212-220.
[26] Sherman, J. and Morrison, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics*, **21**, 124-127.
[27] Tezuka, T. and Kuroki, M. (2022). Formulating the unbiased estimator of the causal effect on the variance with an application to the anomaly detection problem. Under review.
[28] Tian, J. (2004). Identifying linear causal effects. *Proceeding of the 19th National Conference on Artificial Intelligence*, 104-111.
[29] Weiss, N. A., Holmes, P. T. and Hardy, M. (2006). *A course in probability*. Pearson Addison Wesley.
[30] Wright, S. (1923). The theory of path coefficients: A reply to Niles' criticism. *Genetics*, **8**, 239-255.
[31] Wright, S.(1934). The method of path coefficients. *Annals of Mathematical Statistics*, **5**, 161-215.

Table 2. Numerical Experiments.

Case 1: $\sigma_{yy|x} = 0.510$

Sample size	10						25					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	0.509	0.511	0.509	0.544	0.509	0.514	0.510	0.511	0.510	0.522	0.510	0.511
Equation (17)/(21)	0.059	0.052	0.065	0.052	0.059	0.052	0.022	0.021	0.022	0.021	0.022	0.021
var	0.058	0.058	0.064	0.069	0.058	0.058	0.022	0.022	0.023	0.023	0.022	0.022
Skewness	0.937	0.936	1.038	1.000	0.937	0.936	0.570	0.570	0.576	0.566	0.571	0.570
Kurtosis	4.326	4.325	4.908	4.780	4.328	4.327	3.490	3.490	3.481	3.463	3.492	3.492
Minimum	0.016	0.017	0.016	0.018	0.016	0.019	0.102	0.103	0.105	0.108	0.100	0.102
1st Quartile	0.333	0.335	0.324	0.352	0.333	0.338	0.404	0.405	0.402	0.413	0.404	0.406
Median	0.472	0.474	0.467	0.501	0.472	0.477	0.496	0.497	0.495	0.507	0.495	0.497
3rd Quartile	0.644	0.647	0.649	0.691	0.644	0.650	0.601	0.601	0.603	0.615	0.601	0.603
Maximum	2.053	2.056	2.995	3.071	2.052	2.057	1.344	1.345	1.371	1.387	1.344	1.346

Sample size	50						100					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	0.511	0.511	0.511	0.516	0.511	0.511	0.509	0.510	0.509	0.512	0.509	0.510
Equation (17)/(21)	0.011	0.010	0.011	0.010	0.011	0.010	0.005	0.005	0.005	0.005	0.005	0.005
var	0.011	0.011	0.011	0.011	0.011	0.011	0.005	0.005	0.005	0.005	0.005	0.005
Skewness	0.407	0.407	0.411	0.407	0.408	0.407	0.300	0.300	0.301	0.300	0.300	0.300
Kurtosis	3.237	3.237	3.238	3.233	3.239	3.239	3.110	3.110	3.115	3.114	3.110	3.110
Minimum	0.210	0.211	0.205	0.209	0.210	0.211	0.265	0.265	0.266	0.268	0.265	0.265
1st Quartile	0.438	0.438	0.437	0.442	0.438	0.438	0.459	0.459	0.459	0.461	0.459	0.459
Median	0.504	0.504	0.504	0.509	0.504	0.504	0.506	0.506	0.506	0.508	0.506	0.506
3rd Quartile	0.576	0.576	0.576	0.582	0.576	0.577	0.557	0.557	0.557	0.559	0.557	0.557
Maximum	1.034	1.035	1.044	1.053	1.035	1.035	0.867	0.867	0.876	0.880	0.867	0.867

Case 2: $\sigma_{yy|x} = 0.510$

Sample size	10						25					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	0.510	0.603	0.510	0.658	0.511	0.744	0.511	0.541	0.511	0.556	0.511	0.580
Equation (17)/(21)	0.073	0.052	0.107	0.052	0.145	0.052	0.023	0.021	0.024	0.021	0.025	0.021
var	0.073	0.099	0.110	0.155	0.145	0.221	0.023	0.026	0.024	0.028	0.025	0.031
Skewness	1.478	1.531	4.255	4.000	4.826	4.528	0.620	0.617	0.709	0.697	0.748	0.714
Kurtosis	10.112	11.673	87.363	74.588	94.755	74.449	3.688	3.673	3.964	3.932	4.234	4.070
Minimum	-0.814	0.022	-2.621	0.022	-5.144	0.026	0.108	0.113	0.107	0.115	0.107	0.120
1st Quartile	0.318	0.377	0.300	0.400	0.288	0.446	0.403	0.427	0.400	0.436	0.399	0.455
Median	0.464	0.548	0.450	0.585	0.443	0.655	0.496	0.525	0.494	0.538	0.494	0.561
3rd Quartile	0.652	0.769	0.646	0.827	0.648	0.935	0.602	0.637	0.602	0.655	0.602	0.683
Maximum	5.069	6.851	13.617	15.156	13.302	15.291	1.624	1.706	1.653	1.756	1.940	2.073

Sample size	50						100					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	0.510	0.524	0.510	0.531	0.510	0.542	0.510	0.517	0.510	0.520	0.510	0.525
Equation (17)/(21)	0.011	0.010	0.011	0.010	0.011	0.010	0.005	0.005	0.005	0.005	0.005	0.005
var	0.011	0.011	0.011	0.012	0.011	0.012	0.005	0.005	0.005	0.006	0.005	0.006
Skewness	0.390	0.391	0.416	0.414	0.423	0.418	0.300	0.300	0.310	0.309	0.310	0.308
Kurtosis	3.206	3.207	3.270	3.267	3.297	3.288	3.152	3.152	3.187	3.185	3.185	3.183
Minimum	0.193	0.198	0.186	0.196	0.186	0.199	0.247	0.250	0.245	0.250	0.245	0.253
1st Quartile	0.436	0.448	0.435	0.454	0.435	0.463	0.459	0.465	0.458	0.468	0.458	0.472
Median	0.503	0.516	0.502	0.523	0.502	0.534	0.506	0.513	0.506	0.516	0.506	0.522
3rd Quartile	0.576	0.592	0.577	0.601	0.577	0.613	0.557	0.565	0.557	0.568	0.557	0.574
Maximum	0.994	1.026	1.071	1.113	1.110	1.171	0.898	0.910	1.010	1.030	1.009	1.037

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(21): the exact and asymptotic variances derived from equations (17) and (21) with Table 1; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 2. Numerical Experiments.

Case 3:$\sigma_{yy|x}$ = 1.118

Sample size	10						25					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	1.119	1.234	1.119	1.373	1.120	1.409	1.116	1.153	1.118	1.196	1.117	1.202
Equation (17)/(21)	0.397	0.289	0.596	0.331	0.566	0.308	0.128	0.115	0.154	0.132	0.141	0.123
var	0.408	0.457	0.584	0.725	0.595	0.721	0.130	0.135	0.157	0.171	0.145	0.156
Skewness	1.837	1.873	2.391	2.328	2.678	2.774	0.751	0.745	1.040	1.006	0.895	0.867
Kurtosis	14.121	15.044	15.610	14.888	24.310	24.522	4.008	4.012	5.210	5.073	4.491	4.418
Minimum	-0.420	0.043	-2.802	0.043	-8.253	0.051	0.161	0.179	0.168	0.184	0.161	0.191
1st Quartile	0.668	0.758	0.617	0.804	0.620	0.849	0.860	0.892	0.838	0.903	0.846	0.921
Median	0.999	1.108	0.952	1.194	0.964	1.238	1.074	1.110	1.060	1.138	1.065	1.149
3rd Quartile	1.433	1.568	1.421	1.721	1.427	1.756	1.328	1.370	1.333	1.422	1.331	1.425
Maximum	13.817	14.935	12.257	13.943	17.465	18.915	3.792	3.959	4.504	4.658	4.427	4.560

Sample size	50						100					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	1.117	1.134	1.117	1.153	1.117	1.156	1.119	1.127	1.119	1.136	1.119	1.138
Equation (17)/(21)	0.061	0.058	0.071	0.066	0.065	0.062	0.030	0.029	0.034	0.033	0.032	0.031
var	0.061	0.062	0.071	0.074	0.066	0.069	0.030	0.030	0.034	0.035	0.032	0.032
Skewness	0.516	0.513	0.679	0.670	0.611	0.600	0.337	0.336	0.443	0.439	0.392	0.388
Kurtosis	3.461	3.457	3.898	3.881	3.718	3.697	3.184	3.184	3.342	3.337	3.277	3.273
Minimum	0.337	0.344	0.340	0.360	0.332	0.357	0.561	0.567	0.555	0.568	0.575	0.586
1st Quartile	0.943	0.959	0.929	0.962	0.935	0.971	0.998	1.006	0.988	1.005	0.993	1.011
Median	1.096	1.114	1.088	1.125	1.092	1.132	1.110	1.118	1.105	1.123	1.108	1.127
3rd Quartile	1.269	1.288	1.274	1.314	1.272	1.315	1.229	1.238	1.234	1.253	1.231	1.251
Maximum	2.435	2.470	3.173	3.241	2.545	2.609	2.075	2.091	2.111	2.133	2.071	2.092

Case 4:$\sigma_{yy|x}$ = 1.000

Sample size	10						25					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	1.001	1.094	0.999	1.220	0.999	1.233	1.000	1.030	0.999	1.067	1.000	1.069
Equation (17)/(21)	0.316	0.232	0.481	0.270	0.436	0.247	0.102	0.093	0.126	0.108	0.113	0.099
var	0.317	0.349	0.483	0.593	0.463	0.526	0.102	0.105	0.124	0.135	0.113	0.120
Skewness	1.444	1.449	3.284	3.150	0.595	2.589	0.731	0.726	1.016	0.988	0.861	0.840
Kurtosis	7.138	7.186	39.242	34.466	96.540	22.520	3.903	3.899	5.103	5.015	4.351	4.314
Minimum	-2.116	0.038	-3.755	0.040	-29.146	0.051	0.164	0.170	0.161	0.175	0.157	0.176
1st Quartile	0.603	0.678	0.553	0.718	0.563	0.753	0.770	0.797	0.749	0.806	0.760	0.821
Median	0.897	0.985	0.851	1.061	0.868	1.092	0.964	0.993	0.948	1.015	0.955	1.024
3rd Quartile	1.278	1.384	1.270	1.529	1.276	1.539	1.189	1.222	1.193	1.270	1.191	1.268
Maximum	7.246	7.902	18.483	19.766	13.702	15.580	3.042	3.114	4.030	4.203	3.577	3.719

Sample size	50						100					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	1.000	1.015	1.000	1.032	1.001	1.033	1.000	1.007	1.000	1.015	1.000	1.016
Equation (17)/(21)	0.049	0.046	0.058	0.054	0.052	0.049	0.024	0.023	0.028	0.027	0.025	0.025
var	0.049	0.050	0.059	0.061	0.054	0.055	0.024	0.024	0.028	0.029	0.026	0.026
Skewness	0.515	0.512	0.677	0.669	0.601	0.592	0.360	0.359	0.461	0.457	0.412	0.409
Kurtosis	3.500	3.498	3.871	3.854	3.709	3.690	3.240	3.239	3.407	3.402	3.324	3.321
Minimum	0.324	0.330	0.313	0.329	0.312	0.330	0.505	0.510	0.481	0.490	0.488	0.499
1st Quartile	0.844	0.857	0.829	0.857	0.837	0.867	0.892	0.898	0.881	0.895	0.887	0.902
Median	0.981	0.995	0.975	1.006	0.979	1.011	0.991	0.998	0.988	1.003	0.989	1.005
3rd Quartile	1.139	1.154	1.144	1.179	1.141	1.176	1.099	1.107	1.105	1.122	1.102	1.118
Maximum	2.508	2.532	2.559	2.590	2.419	2.447	1.803	1.814	2.020	2.044	1.877	1.895

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(21): the exact and asymptotic variances derived from equations (17) and (21) with Table 1; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 2. Numerical Results.

Case 5: $\sigma_{yy|x} = 6.890$

Sample size	10						25					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	6.889	8.935	6.846	11.998	6.818	11.758	6.891	7.478	6.839	8.310	6.849	8.093
Equation (17)/(20)	89.284	39.921	273.079	85.493	258.651	69.743	20.461	15.968	48.386	34.197	39.017	27.897
var	89.195	96.603	269.086	288.518	261.860	276.373	20.314	20.997	48.615	50.410	39.916	41.375
Skewness	3.174	3.590	3.751	4.816	3.909	5.284	1.404	1.435	1.944	2.010	1.797	1.864
Kurtosis	27.518	32.126	45.247	57.833	50.464	69.849	6.565	6.677	10.115	10.369	9.068	9.359
Minimum	-41.487	0.043	-217.754	0.057	-209.358	0.047	-1.922	0.227	-7.491	0.217	-6.815	0.241
1st Quartile	1.233	2.735	-1.117	2.227	-0.831	2.314	3.679	4.201	1.922	3.255	2.389	3.519
Median	4.427	6.001	2.666	6.404	2.911	6.363	5.996	6.543	5.160	6.467	5.372	6.531
3rd Quartile	9.754	11.656	10.415	14.959	10.489	14.655	9.120	9.719	9.769	11.290	9.641	10.861
Maximum	227.310	249.270	458.364	521.416	428.839	538.489	48.157	49.007	87.847	91.660	85.278	88.653

Sample size	50						100					
Variables	Z_1		Z_2		$\{Z_1, Z_2\}$		Z_1		Z_2		$\{Z_1, Z_2\}$	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	6.907	7.176	6.916	7.589	6.910	7.466	6.886	7.015	6.879	7.200	6.889	7.152
Equation (17)/(20)	8.971	7.984	20.098	17.099	16.240	13.949	4.225	3.992	9.245	8.549	7.501	6.974
var	9.001	9.149	20.271	20.664	16.558	16.865	4.206	4.241	9.205	9.290	7.575	7.643
Skewness	0.916	0.925	1.248	1.270	1.137	1.156	0.615	0.617	0.820	0.825	0.767	0.772
Kurtosis	4.414	4.427	5.743	5.808	5.190	5.245	3.591	3.593	4.037	4.042	3.878	3.882
Minimum	0.162	0.406	-1.882	0.327	-1.062	0.364	1.473	1.617	0.203	0.443	0.326	0.613
1st Quartile	4.758	5.004	3.666	4.310	3.981	4.494	5.426	5.546	4.696	4.999	4.906	5.156
Median	6.487	6.751	6.093	6.739	6.213	6.751	6.674	6.801	6.467	6.781	6.547	6.802
3rd Quartile	8.574	8.852	9.264	9.949	9.050	9.623	8.135	8.269	8.616	8.947	8.497	8.768
Maximum	27.249	27.699	44.779	46.374	39.581	40.831	19.980	20.117	25.766	26.199	22.501	22.832

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(21): the exact and asymptotic variances derived from equations (17) and (21) with Table 1; Var: empirical variances from 50000 estimated causal effects on the variance.