An Unbiased Estimator of the Causal Effect on the Variance based on the Back-door Criterion 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 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) of the outcome variable with 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

1.1. Backgound

Statistical causal inference using linear structural equation models (linear SEMs) has been widely used to clarify cause-effect relationships between random variables in fields such as sociology, economics, and biology, 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 important aspects are direct, indirect, and total effects (Bollen, 1989). According to Bollen (1987, p.40), intuitively, 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." Here, an “intervening variable" is a random variable that could be affected by a treatment variable and have an effect on an outcome 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 due to external intervention. The causal understanding regarding the difference of total, direct and indirect effects contributes to evaluating how much of the causal effect of a treatment variable on an outcome variable is captured/ not captured by intervening variables. The statistical method for promoting such causal understanding is called mediation analysis, which has its roots in the literature of linear SEMs, going back to path analysis (Wright, 1923, 1934) and continuing in the social sciences through the works of Duncan (1975), Baron and Kenny (1986) and Bollen (1989).

To evaluate the total effect, which this paper focuses on, 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;

*Corresponding author. Email address:teduka-taiki-bs@ynu.jp

Perković, 2018; Tian, 2004). Herein, “identifiable" indicates that the total effect can be uniquely determined based on the variance-covariance parameters of observed variables.

When we wish to characterize the distributional change introduced by 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. 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 external intervention (the causal effect on the mean). However, it is often necessary to evaluate the variation (variance) of the outcome variable due to 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 due to external intervention, thereby reducing the variation (or minimizing 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), when treating hyperglycemia, 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 physician’s knowledge, plays an important role in constructing the acceptable range to detect a threat to a patient’s health.

Regarding the estimation accuracy (or, the variance) of the causal effect on the variance, when the ordinary least squares 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 intervention 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 external intervention. Here, it is noted that the existing estimators of the causal effect on the variance are the consistent but not unbiased estimators. Estimation accuracy problems are essential issues related to statistical causal inference, and thus it is important to formulate the unbiased estimator of the causal effect on the variance with the excat variance. This is because the reliable evaluation of estimation accuracy of the causal effect on 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 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 ordinary least squares 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 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.

1.2. Motivating Example

To motivate our problem, consider a case study of setting up coating conditions for car bodies, reported by Okuno et al. (1986). According to Okuno et al. (1986), since car bodies are coated in order to increase both the rust protection quality and the visual appearance, a certain level of the coating thickness must be ensured in the coating process. At that time, the coating process was conducted by operators who sprayed the car bodies with the paint. This was dependent on operators' skills and might cause low transfer efficiency. Okuno et al. (1986) collected non-experimental data in the coating process in order to examine the process conditions and to increase the transfer efficiency, which were important to establish the automated stable manufacturing process. The sample size is 38 and the observed variables of interest are the following:

Coating Conditions: Dilution ratio (X_1), Degree of viscosity (X_2), Temperature of the paints (X_8)

Spraying Conditions: Gun speed (X_3), Spray-distance (X_4), Air pressure (X_5), Pattern width (X_6), Fluid output (X_7)

Environment Conditions: Temperature (X_9), Degree of moisture (X_{10})

Response: Transfer efficiency (Y), which was defined as "the coated paint volume"/"the consumption of paints"$\times 100\%$

According to Okuno et al. (1986), dilution ratio (X_1) and spray-distance (X_4) are easy to be controlled. Degree of viscosity (X_2), gun speed (X_3), air pressure (X_5) and pattern width (X_6) are able to be controlled to some extent. Fluid output (X_7) and temperature of the paints (X_8) are results from other factors and are difficult to be controlled. Temperature (X_9) and degree of moisture (X_{10}) are environment conditions that cannot be controlled. In addition, Okuno et al. (1986) also considered "wind speed" (environment condition), "solid content" (coating condition) and others as factors which might have an effect on Transfer efficiency (Y). However, these factors were not observed, because it seems to be sufficient to observe the ten variables above to achieve their aim, according to Okuno et al. (1986).

Concerning the coating process, Okuno et al. (1986) provided the sample correlation matrix shown in Table 3. By applying conventional stepwise regression analysis to Table 3 according to Okuno et al. (1986), the following regression model is obtained:

$$Y = -0.636x_4 - 0.465x_6 + 0.189x_7 - 0.372x_8. \quad (1)$$

It is seen from the regression model (1) that the transfer efficiency (Y) can be increased by controlling X_4, X_6, X_7 and X_8 according to Okuno et al. (1986), but note that both the fluid output (X_7) and the temparature of the paints (X_8) are difficult to be controlled actually. In addition, in order to establish stable manufacturing process, it is important to understand how the variation of the transfer efficiency (Y) would change by external intervention, because the increase in the variation of the transfer efficiency (Y) may lead to the construction of the unstable coating process. However, from equation (1), it is difficult to understand how the variation in the transfer efficiency (Y) would change by external intervention: the analysis should not be simply based on statistical aspects, but it is desirable to describe the cause-effect relationships as a directed graph (which is called a causal path diagram) according to the analyst's knowledge. Then, combining the causal knowledge with statistical data, statistical causal inference using linear SEMs enables us to evaluate the variation of the transfer efficiency (Y) due to external intervention through (non-experimental) statistical data collected from the current coating process.

Here, to present our results, according to Kuroki (2008, 2012), assume that the cause-effect relationships in the coating process are given in Figure 1. For example, intuitively, in Figure 1, a directed edge from X_1 to X_2 ($X_1 \rightarrow X_2$) means that X_1 could cause X_2 directly, and a directed path from X_1 to X_7 with a missing directed edge ($X_1 \rightarrow X_2 \rightarrow X_7$) means that the effect of X_1 on X_7 could be mediated by X_2 but can not directly. Here, this paper will not discuss statistical inference problem of Figure 1. Refer to Kuroki (2012) for details on this case study.

Here, under the assumption that $X_1, X_2, ..., X_{10}, Y$ follows the multivariate normal distribution with zero mean vector and the variance-covariance matrix shown in Table 3, we evaluated unbiased estimators (17) and the consistent estimators (22) of the causal effect on the variance of the transfer efficiency (Y) 5000 times based on the sample size 38. Table 3 reports the basic statistics of the unbiased estimators (17) and the consistent estimators (22) when a set of variables given in 'Variables' rows are utilized to identify the causal effects.

Regarding the causal effect on the variance, from the "Estimates" rows of Table 2, although the consistent estimators are different from the values of equation (11), the unbiased estimators are close to the values of equation (11) even for the small sample sizes ($n = 38$). Especially, regarding the external intervention to dilution ratio (X_1), the unbiased estimators show that the external intervention could reduce the variation of transfer efficiency (Y), but the consistent estimators imply that the external intervention does not reduce the variation of transfer efficiency (Y). Such difference may lead to the serious practical judgments: to establish the stable manufacturing process increase the transfer efficiency,

Table 1: The sample correlation matrix (Okuno et al. 1986)

	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	-0.678	-0.215	0.230	0.040	0.116	0.338	0.002	0.145	-0.496	-0.198
X_2	-0.678	1.000	0.241	-0.442	-0.024	0.005	-0.422	-0.590	-0.509	0.684	0.463
X_3	-0.215	0.241	1.000	-0.201	0.004	-0.067	0.208	-0.007	-0.082	0.307	0.292
X_4	0.230	-0.442	-0.201	1.000	0.191	-0.286	0.287	0.446	0.521	-0.477	-0.614
X_5	0.040	-0.024	0.004	0.191	1.000	0.291	0.117	0.034	-0.048	0.010	-0.151
X_6	0.116	0.005	-0.067	-0.286	0.291	1.000	0.057	-0.123	-0.147	0.178	-0.226
X_7	0.338	-0.422	0.208	0.287	0.117	0.057	1.000	0.251	0.287	-0.122	-0.113
X_8	0.002	-0.590	-0.007	0.446	0.034	-0.123	0.251	1.000	0.761	-0.342	-0.551
X_9	0.145	-0.509	-0.082	0.521	-0.048	-0.147	0.287	0.761	1.000	-0.571	-0.431
X_{10}	-0.496	0.684	0.307	-0.477	0.010	0.178	-0.122	-0.342	-0.571	1.000	0.282
Y	-0.198	0.463	0.292	-0.614	-0.151	-0.226	-0.113	-0.551	-0.431	0.282	1.000

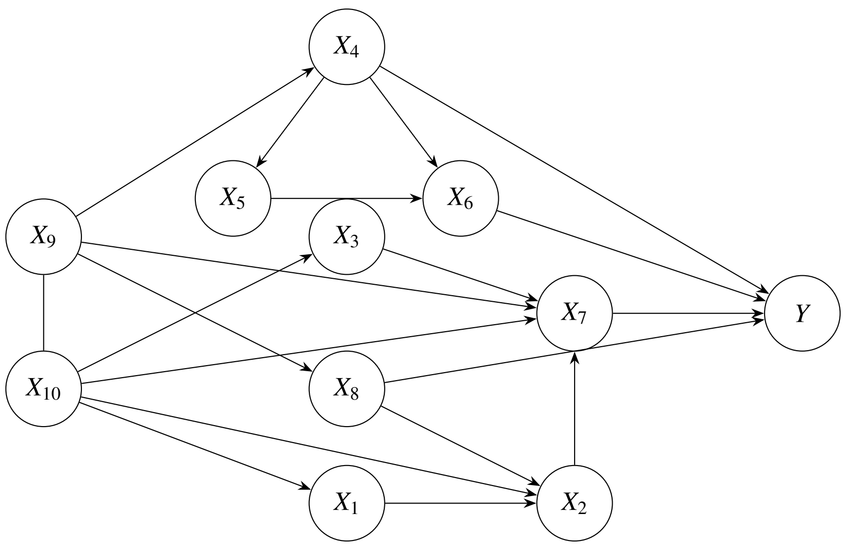

Figure 1: Causal path diagram of the coating process (Kuroki, 2012)

Table 2: Basis statistics of the coating process

Treatment variable	Dilution Ratio (X_1)				Spray Distance (X_4)					
Covariate	X_{10}		$\{X_9, X_{10}\}$		X_9		$\{X_7, X_8, X_9\}$		$\{X_7, X_8\}$	
Equation(11)	0.975		0.963		0.629		0.631		0.640	
Estimator	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent	unbiased	consistent
Estimates	0.989	1.022	0.972	1.023	0.627	0.649	0.632	0.679	0.636	0.667
Equation(17)/(22)	0.055	0.052	0.052	0.049	0.023	0.021	0.023	0.022	0.024	0.023
var	0.056	0.060	0.053	0.058	0.022	0.023	0.023	0.026	0.023	0.024
Skewness	0.576	0.575	0.377	0.372	0.515	0.513	0.502	0.495	0.514	0.505
Kurtosis	3.722	3.724	3.034	3.038	3.453	3.450	3.429	3.411	3.535	3.513
Minimum	0.366	0.378	0.303	0.318	0.186	0.194	0.247	0.265	0.216	0.227
1st Quartile	0.822	0.851	0.808	0.852	0.521	0.540	0.525	0.565	0.531	0.558
Median	0.969	1.001	0.954	1.004	0.614	0.634	0.623	0.669	0.623	0.655
3rd Quartile	1.130	1.167	1.122	1.179	0.719	0.744	0.727	0.780	0.729	0.764
Maimimum	2.285	2.351	1.878	1.945	1.278	1.320	1.340	1.412	1.395	1.460

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; Equation(11): the causal effect on the variance from equation (11) with Table 1; Equation(17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 1; Var: empirical variances from 50000 estimated causal effects on the variance.

the external intervention should be conducted from the viewpoint of the unbiased estimators, but not from the viewpoint of the consistent estimators.

2. Preliminaries

2.1. 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 with the length m 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 with the length m is a sequence $a = a_0, a_1, \cdots, b = a_m$ of distinct vertices such that $a_{i-1} \rightarrow a_i$ 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. Especially, $(a, b) \in E$ for $a, b \in \boldsymbol{V}$ is a directed edge from a to b and the directed path from a to b with the length 1 at the same time. a is a parent of b and an ancestor of b at the same time. b is a child of a and a descendant of b at the same time.

When the set of descendants of a is denoted as $de(a)$, the vertices in $V \backslash (de(a) \cup \{a\})$ are said to be the nondescendants of a. A vertex is said to be a collider if it is a common child of the other two or more vertices; otherwise, it is said to be a non-collider. 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).

2.2. 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). 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 $\boldsymbol{V} = \{V_1, V_2, \cdots, V_p\}$ of p continuous random variables and a set $\boldsymbol{E}$ of directed edges are given. Then, the DAG G is called the causal path diagram if the

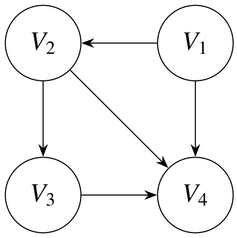

Figure 2: Causal path diagram

random variables are generated by a Gaussian linear SEM

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

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 $\mathbf{0}_p$ be an p-dimensional vector whose i-th element is zero for $i = 1, 2, ..., p$, $\boldsymbol{\epsilon}_v = (\epsilon_{v_1}, \epsilon_{v_2}, \dots, \epsilon_{v_p})$ denotes a set of random variables, which is assumed to follow the multivariate normal distribution with the mean vector $\mathbf{0}_p$ 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 $i, j = 1, 2, ..., p$ $(i \neq j)$ are referred to as the intercept of V_i and the causal path coefficient (or direct effect) of V_j on V_i, respectively. □

As an example, consider the causal path diagram shown in Figure 2. From Figure 2, we can judge that: (1) V_1 could be a direct cause of V_2 and V_4, (2) V_2 could be a direct cause of V_3 and V_4, and (3) V_3 could be a direct cause of V_4. Then, the Gaussian linear SEM defined by Figure 2 is as

$$V_1 = \alpha_{v_1} + \epsilon_{v_1}, \quad V_2 = \alpha_{v_2} + \alpha_{v_2 v_1} V_1 + \epsilon_{v_2}, \quad V_3 = \alpha_{v_3} + \alpha_{v_3 v_2} V_2 + \epsilon_{v_3}, \quad V_4 = \alpha_{v_4} + \alpha_{v_4 v_1} V_1 + \alpha_{v_4 v_2} V_2 + \alpha_{v_4 v_3} V_3 + \epsilon_{v_4}, \tag{3}$$

where $\epsilon_{v_1}, \epsilon_{v_2}, \epsilon_{v_3}, \epsilon_{v_4}$ follow the normal distribution with zero mean and non-zero variance independently.

The conditional independence induced by the Gaussian linear SEM (2) 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 $\boldsymbol{Z}$. □

In Figure 2, both $\{V_2\}$ and $\{V_2, V_4\}$ satisfy Condition (1) of Definition 2 on the path $V_1 \to V_2 \to V_3$ since both sets include a non-collider V_2. However, a collider (V_4) on the other paths is in $\{V_2, V_4\}$ but not in $\{V_2\}$. Thus, V_2 d-separates V_1 from V_3 but $\{V_2, V_4\}$ does not.

If $\boldsymbol{Z}$ d-separates X from Y in the causal path diagram G, then X is conditionally independent of Y given $\boldsymbol{Z}$ in the corresponding linear SEM (e.g., Pearl, 2009). For example, in Figure 2, since $\{V_2\}$ d-separates $\{V_1\}$ from $\{V_3\}$ respectively, V_1 is conditionally independent of V_3 given V_2.

2.3. 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 (2), 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 (2).

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

$$f(y|\mathrm{do}(X = x)) = \int_w \frac{f(x, y, \boldsymbol{w})}{f(x|\mathrm{pa}(x))} d\boldsymbol{w} \tag{4}$$

(Pearl, 2009). When equation (4) 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,

$$E(Y|\mathrm{do}(X=x)) = \mu_{y|x} = \int_y yf(y|\mathrm{do}(X=x))\,dy, \quad \mathrm{var}(Y|\mathrm{do}(X=x)) = \sigma_{yy|x} = \int_y (y-\mu_{y|x})^2 f(y|\mathrm{do}(X=x))\,dy \tag{5}$$

are called the causal effect of $\mathrm{do}(X=x)$ on the mean of Y and the causal effect of $\mathrm{do}(X=x)$ on the variance of Y, 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 (2), the first derivative of $E(Y|\mathrm{do}(X=x))$ of Y, namely,

$$\frac{dE(Y|\mathrm{do}(X=x))}{dx} = \tau_{yx} \tag{6}$$

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 (2) 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 $G_{\underline{X}}$. □

Regarding other identification conditions of causal effects, for example, "the front door criterion" (Pearl, 2009) and “the effect restoration" (Kuroki and Pearl, 2014) are known. However, this paper is only concerned with identification of a causal effect using the back door criterion. As seen from the description of Definition 3, the back-door criterion is not a statistical concept, and can not be tested through statistical data.

In Figure 2, both $\{V_2\}$ and $\{V_1, V_2\}$ satisfy the back door criterion relative to (V_3, V_4). However, $\{V_1\}$ does not satisfy the back door criterion relative to (V_3, V_4), since $\{V_1\}$ does not d-separate V_3 from V_4 in the graph $G_{\underline{V}_3}$ derived from Figure 2. For example, $\{V_1\}$ does not include any non-collider (V_2) on the path $V_3 \leftarrow V_2 \rightarrow V_4$ and neither colliders nor their descendants are not on the path.

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

$$f(y|\mathrm{do}(X=x)) = \int_z f(y|x,z)f(z)dz \tag{7}$$

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

$$Y = \beta_{y.xz} + \beta_{yx.xz}X + B_{yz.xz}\boldsymbol{Z} + \epsilon_{y.xz}, \tag{8}$$

where $\epsilon_{y.xz}$ is a random variable of the regression model (8) that has a normal distribution with mean zero 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 (8), respectively. Here, according to the standard assumption of linear regression analysis, in the regression model (8), $\epsilon_{y.xz}$ is assumed to be independent of both X and $\boldsymbol{Z}$. Then, for a non-empty set $\boldsymbol{Z}$, letting

$$\sigma_{xy\cdot z} = \sigma_{xy} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zy}, \quad \sigma_{xx.z} = \sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}, \quad \Sigma_{zz.x} = \Sigma_{zz} - \frac{\Sigma_{zx}\Sigma_{xz}}{\sigma_{xx}}, \quad \Sigma_{yz.x} = \Sigma_{yz} - \frac{\sigma_{xy}}{\sigma_{xx}}\Sigma_{xz}, \quad \Sigma_{zy.x} = \Sigma'_{yz.x}, \tag{9}$$

the regression coefficient of X and the regression coefficient vector of $\mathbf{Z}$ 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 $\sigma_{xx} \neq 0$, $\sigma_{xx.z} \neq 0$, and both Σ_{zz} and $\Sigma_{zz.x}$ are positive definite matrices.

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 (7), consider the regression model of Y on X and $\mathbf{Z}$, namely, equation (8). Then, letting $\sigma_{yy.z} = \sigma_{yy} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zy}$ and $\sigma_{yy.xz} = \sigma_{yy.z} - \frac{\sigma_{xy.z}^2}{\sigma_{xx.z}}$, $E(Y|\mathrm{do}(X = x))$ and $\mathrm{var}\,(Y|\mathrm{do}(X = x))$ are formulated as

$$E(Y|\mathrm{do}(X = x)) = \mu_{y|x} = \mu_y + \beta_{yx.xz}(x - \mu_x) = \mu_y + \tau_{yx}(x - \mu_x) \tag{10}$$

and

$$\mathrm{var}\,(Y|\mathrm{do}(X = x)) = \sigma_{yy|x} = \sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz}, \tag{11}$$

respectively (Kuroki and Miyakawa, 1999ab, 2003). Here, equation (11) shows that $\mathbf{Z}$ behaves similarly to the random variable such as $\epsilon_{y.xz}$ in equation (8) by conducting the external intervention $\mathrm{do}(X = x)$, and the external intervention may not reduce the variation of the outcome variable Y (Kuroki, 2012).

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

$$X = \beta_{x.z} + B_{xz.z}\mathbf{Z} + \epsilon_{x.z}, \tag{12}$$

where $\epsilon_{x.z}$ is a random variable of the regression model (12) that has a normal distribution with mean zero 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 (12), respectively. Here, $\epsilon_{x.z}$ 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. Then, for non-empty set $\mathbf{Z}$, letting

$$s_{xy.z} = s_{xy} - S_{xz}S_{zz}^{-1}S_{zy}, \quad s_{xx.z} = s_{xx} - S_{xz}S_{zz}^{-1}S_{zx}, \quad S_{zz.x} = S_{zz} - \frac{S_{zx}S_{xz}}{s_{xx}}, \quad S_{yz.x} = S_{yz} - \frac{s_{xy}}{s_{xx}}S_{xz}, \quad S_{zy.x} = S'_{yz.x}, \tag{13}$$

through the ordinary least squares method, the unbiased estimators of $\beta_{yx.xz}$, $B_{xz.z}$ and $B_{yz.xz}$ of equations (8) and (12) 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 $s_{xx} \neq 0$, $s_{xx.z} \neq 0$ and both S_{zz} and $S_{zz.x}$ are positive definite matrices. Here, letting n and q be the sample size and the number of random variables in $\mathbf{Z}$, respectively, for $q < n - 2$,

$$\hat{\sigma}_{yy.xz} = \frac{s_{yy.xz}}{n - q - 2} = \frac{s_{yy.z} - \frac{s_{xy.z}^2}{s_{xx.z}}}{n - q - 2}, \quad \hat{\Sigma}_{zz} = \frac{1}{n - 1}S_{zz} \tag{14}$$

are also unbiased estimators of $\sigma_{yy.xz}$ and Σ_{zz}, respectively, where $s_{yy.z} = s_{yy} - S_{yz}S_{zz}^{-1}S_{zy}$.

Under the random sampling, when the total effect τ_{yx} is estimated as $\hat{\tau}_{yx} = \hat{\beta}_{yx\cdot xz}$ through the ordinary least squares method in the regression model (8), the exact variance of $\hat{\beta}_{yx\cdot xz}$ is given by

$$\mathrm{var}\left(\hat{\beta}_{yx\cdot xz}\right) = \frac{1}{n - q - 3}\frac{\sigma_{yy\cdot xz}}{\sigma_{xx\cdot z}} \tag{15}$$

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

The following theorem holds:

Theorem 1. *Under the Gaussian linear SEM (2), suppose that* $\mathbf{Z}$ *satisfies the back-door criterion relative to* (X, Y) *in the causal path diagram G. When the ordinary least squares method is utilized to evaluate the statistical parameters in equations (10) and (11), the unbiased estimators of* $\mu_{y|x} = E(Y|do(X = x))$ *and* $\sigma_{yy|x} = var(Y|do(X = x))$ *are given by*

$$\hat{\mu}_{y|x} = \hat{\mu}_y + \hat{\beta}_{yx.xz}(x - \hat{\mu}_x) \tag{16}$$

$$\hat{\sigma}_{yy|x} = \hat{\sigma}_{yy.xz}\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right) + \hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}, \tag{17}$$

respectively. $\hat{\mu}_{y|x}$ *and* $\hat{\sigma}_{yy|x}$ *are called the estimated causal effect of* $do(X = x)$ *on the mean of Y and estimated causal effect of* $do(X = x)$ *on the variance of Y, 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*

$$var\left(\hat{\mu}_{y|x}\right) = \frac{1}{n}\left(\sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz}\right) + \frac{\sigma_{yy.xz}}{(n-q-3)\sigma_{xx.z}}\left((x-\mu_x)^2 + \frac{\sigma_{xx}}{n}\right), \tag{18}$$

$$\begin{aligned}
var(\hat{\sigma}_{yy|x}) = {} & \frac{2(B_{yz.xz}\Sigma_{zz}B'_{yz.xz})^2}{n-1} + \frac{2\sigma^2{}_{yy.xz}}{n-q-2}\left(\left(1-\frac{q}{n-1}\right)^2 - 2\left(1-\frac{q}{n-1}\right)\frac{q\sigma_{xx.z} + (n-1)B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-1)(n-q-3)\sigma_{xx.z}}\right. \\
& \left. +E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{(n-1)s_{xx.z}}\right)^2\right)\right) + \frac{2\sigma^2_{yy.xz}}{(n-1)^2}\left(q + 2\frac{q\sigma_{xx.z} + (n-1)B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)\sigma_{xx.z}} + E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right)\right) \\
& +\frac{4\sigma_{yy.xz}}{(n-1)^2}\left((n-1)B_{yz.xz}\Sigma_{zz}B'_{yz.xz} + E\left(\frac{(B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2}{s_{xx.z}}\right)\right),
\end{aligned} \tag{19}$$

respectively, where

$$\begin{aligned}
E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right) = {} & \frac{2q\sigma^2_{xx.z} + 4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B_{xz.z}}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} + \frac{2(n-1)(B_{xz.z}\Sigma_{zz}B_{xz.z})^2}{(n-q-3)(n-q-5)\sigma^2_{xx.z}} \\
& +\frac{(q\sigma_{xx.z} + (n-1)B_{xz.z}\Sigma_{zz}B_{xz.z})^2}{(n-q-3)(n-q-5)\sigma^2_{xx.z}}
\end{aligned} \tag{20}$$

$$E\left(\frac{(B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2}{s_{xx.z}}\right) = \frac{(n-1)(n(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2 + (B_{xz.z}\Sigma_{zz}B'_{xz.z})(B_{yz.xz}B'_{xz.z})^2 + \sigma_{xx.z}B_{yz.xz}\Sigma_{zz}B'_{yz.xz})}{(n-q-3)\sigma_{xx.z}}. \tag{21}$$

□

Both equations (16) and (18) are given by Kuroki and Nanmo (2020). The derivation of equations (17) and (19), which are the new results, is provided in Appendix. Here, from Appendix, note that the assumption of Gaussian random variables in equation (2) is not necessary to derive equations (17), but necessary to derive equation (19).

For a large sample size n such as $n^{-1} >> n^{-2} \approx 0$, the consistent estimator $\tilde{\sigma}_{yy|x}$ of $\sigma_{yy|x}$ can be given by

$$\tilde{\sigma}_{yy|x} = \hat{\sigma}_{yy.xz} + \hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}, \tag{22}$$

which shows that equation (22) is larger than equation (17). In addition, the asymptotic variance of $\hat{\sigma}_{yy|x}$, a.var$(\hat{\sigma}_{yy|x})$, is given by

$$\begin{aligned}
\text{a.var}(\hat{\sigma}_{yy|x}) &= \frac{2\sigma^2_{yy.xz}}{n} + \frac{2(B_{yz.xz}\Sigma_{zz}B_{yz.xz})^2}{n} + \frac{4\sigma_{yy.xz}}{n}\left(B_{yz.xz}\Sigma_{zz}B'_{yz.xz} + \frac{(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2}{\sigma_{xx.z}}\right) \\
&= \frac{2}{n}\left(\sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz}\right)^2 + \frac{4\sigma_{yy.xz}}{n\sigma_{xx.z}}(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2.
\end{aligned} \tag{23}$$

Here, when we let $\beta_{yx.x} = \sigma_{xy}/\sigma_{xx}$, from $\beta_{yx.x} = \tau_{yx}$ and $B_{xz.z} = \Sigma_{xz}\Sigma_{zz}^{-1}$, the covariance between X and equation (8) leads to

$$\sigma_{xy} = \beta_{yx.xz}\sigma_{xx} + B_{yz.xz}\Sigma_{zx} = \tau_{yx}\sigma_{xx} + B_{yz.xz}\Sigma_{zx}, \tag{24}$$

which provides

$$B_{yz.xz}\Sigma_{zz}B'_{xz.z} = B_{yz.xz}\Sigma_{zx} = (\beta_{yx.x} - \tau_{yx})\sigma_{xx} \tag{25}$$

and

$$\sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz} = \sigma_{yy.x} - B_{yz.xz}\Sigma_{zz.x}B'_{yz.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz} = \sigma_{yy.x} + \frac{(B_{yz.xz}\Sigma_{zx})^2}{\sigma_{xx}} = \sigma_{yy.x} + (\beta_{yx.x} - \tau_{yx})^2\sigma_{xx}. \tag{26}$$

From equation (26), the first term of equation (22), which is equivalent to equation (11), 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, from equation (25), $B_{yz.xz}\Sigma_{zz}B'_{xz.z}$ in the second term of equation (23) 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 (22). From this consideration, letting $\hat{\sigma}_{yy|x,z}$ be the estimated causal effect of do($X = x$) on the variance of Y to emphasize that $\mathbf{Z}$ is utilized to estimate equation (11), 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 (2), 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 ordinary least squares method is utilized to evaluate the statistical parameters in equations (10) and (11), if $\mathbf{Z}_2$ d-separates X from $\mathbf{Z}_1$, then*

$$a.var(\hat{\sigma}_{yy|x.z_1,z_2}) \leq a.var(\hat{\sigma}_{yy|x.z_2}) \tag{27}$$

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

$$a.var(\hat{\sigma}_{yy|x.z_1}) \leq a.var(\hat{\sigma}_{yy|x.z_1,z_2}) \tag{28}$$

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*

$$\frac{\sigma_{yy\cdot xz_1z_2}}{\sigma_{xx\cdot z_1z_2}} \leq \frac{\sigma_{yy\cdot xz_2}}{\sigma_{xx\cdot z_2}} \tag{29}$$

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

$$\frac{\sigma_{yy\cdot xz_1}}{\sigma_{xx\cdot z_2}} \leq \frac{\sigma_{yy\cdot xz_1z_2}}{\sigma_{xx\cdot z_1z_2}} \tag{30}$$

holds.

Intuitively, equation (27) shows that the estimation accuracy could be improved by adding $\mathbf{Z}_1$, because $\mathbf{Z}_1$ is not correlated with X given $\mathbf{Z}_2$ and plays a role in decreasing the residual variance of Y. In contrast, equation (28) shows that the estimation accuracy could be worse, because adding $\mathbf{Z}_2$ may cause the multicollinearity and increases the residual variance of Y. In Figure 1, since V_2 d-separates V_1 from V_3, from Theorem 2, we know

$$\mathrm{a.var}(\hat{\sigma}_{v_4v_4|v_3.v_1v_2}) \leq \mathrm{a.var}(\hat{\sigma}_{v_4v_4|v_3.v_2}) \tag{31}$$

holds from the graph structure without statistical data.

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 $n = 10, 25, 50$, 100, 500 and 1000. For simplicity, consider the DAG depicted in Figure 3 and the Gaussian linear SEM in the form of

$$Y = \alpha_{yx}X + \alpha_{yz_1}Z_1 + \epsilon_y, \quad X = \alpha_{xz_2}Z_2 + \epsilon_x, \quad Z_1 = \alpha_{z_1z_2}Z_2 + \epsilon_{z_1}, \quad Z_2 = \epsilon_{z_2}, \tag{32}$$

where we assume the following two cases as the distribution with mean zero of ϵ_x, ϵ_y, ϵ_{z_1} and ϵ_{z_2} independently; (a) a normal distribution, and (b) a uniform distribution. The matrices of the causal path coefficients of X, Y, Z_1, and Z_2 shown in Table 3 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 also 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 this 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, this 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 in Case 3, whereas $\beta_{yx.x} \neq 0$ and $\tau_{yx} = \beta_{yx.xz} \simeq 0$ hold in Case 4. Case 5 represents an extreme situation in which the simple regression model of Y on X,

$$E(Y|X = x) = \mu_y + \beta_{yx.x}(x - \mu_x),$$

is orthogonal to the causal effect on the mean

$$E(Y|\mathrm{do}(X = x)) = \mu_y + \beta_{yx.xz}(x - \mu_x),$$

i.e., $\beta_{yx.xz}\beta_{yx.x} = \tau_{yx}\beta_{yx.x} \simeq -1$ holds.

We simulated n random samples from a multivariate normal distribution of (X, Y, Z_1, Z_2) with a zero mean vector and the correlation matrices generated from each case of Table 3. Then, regarding the causal effects on the variance, we evaluated both the unbiased estimator (17) and the consistent estimator (22) 50000 times based on $n = 10, 25, 50$, 100, 500 and 1000. Tables 4 and 5 report the basic statistics of equations (17) and (22) when $\{Z_1\}$, $\{Z_2\}$ and $\{Z_1, Z_2\}$ are utilized to identify the causal effects.

First, from the "Estimates" rows of Tables 4 and 5, for each case, the consistent estimators are highly biased in the smaller sample sizes but become less biased in the larger sample sizes. Especially, the bias reduction speed based on the sample size depends on the correlation between X and $\mathbf{Z}$: it seems that it is slower when X is highly correlated with $\mathbf{Z}$. In contrast, the unbiased estimators are close to the true values even for the small sample sizes. However, as seen from the "Minimum" rows of Tables 4 and 5, when X is correlated with $\mathbf{Z}$, the minimum values of the unbiased estimators are negative for the smaller sample size, but not for the larger sample size; the consistent estimators do not take negative values. In addition, from both the "Minimum" and "Maximum" rows of Tables 4 and 5, when X is highly correlated with $\mathbf{Z}$, the sample ranges of the unbiased estimators are wider than those of the consistent estimators in the smaller sample sizes. However, they become close to those of the consistent estimators in the larger sample sizes. Here, note that the sample ranges of the unbiased estimators are narrower than or close to those of the consistent estimators when X is uncorrelated with $\mathbf{Z}$.

Second, from the "(17)/(22)" rows of Tables 4 and 5, except for Case 2, for all sample sizes, equations (17) and (22) when Z_2 is selected are larger than when either Z_1 or $[Z_1, Z_2]$ are selected. Also, equations (17) and (22) when $[Z_1, Z_2]$ is selected are larger than when Z_1 is selected. This implies that the relationships are consistent with the results obtained by Theorem 2. In contrast, in Case 2 with the sample size $n \leq 25$, 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 not only on the sample size but also on the multicollinearity between X and $\mathbf{Z}$ and the number of random variables included in $\mathbf{Z}$: Theorem 2 holds for large sample sizes even 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 when ϵ_x, ϵ_y, ϵ_{z_1} and ϵ_{z_2} follow the normal distribution. In contrast, when ϵ_x, ϵ_y, ϵ_{z_1} and ϵ_{z_2} follow the uniform distribution, the differences between equation (17) and the empirical variances of the unbiased estimator are more significant for the smaller sample size, but are not for the larger sample size. In addition, when X is correlated with $\boldsymbol{Z}$, the asymptotic variance (22) 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 (22) and the empirical variances of the consistent estimator are significant when X is correlated with $\boldsymbol{Z}$. However, the differences between the variables becomes smaller 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 $\boldsymbol{Z}$, both unbiased and consistent estimators take large values in the small sample size, which implies that these estimators are unstable under 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, when the ordinary least squares method is utilized to estimate the total effect, 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) 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. To avoid the problem, we showed in Theorem 1 and numerical experiments that the variance estimator, equation (17), performs better than equation (22) in small samples. Theorem 1 would help statistical practitioners to predict appropriately what would happen to the variation of the outcome variable when conducting external intervention. In addition, Theorem 2 shows that the asymptotic estimation accuracy of the estimated causal effect on the variance depends on the selection of random variables that satisfies the back door criterion, and there are some situation where such a difference can be read-off from the graph structure, before sampling statistical data.

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 variables that are not affected by the treatment variable. 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, when the statistical causal model is not consistent with available data. One of our suggestions to solve the problem is to use $\max\{0, \hat{\sigma}_{yy|x}\}$ instead of $\hat{\sigma}_{yy|x}$ to evaluate the causal effect on the variance. However, $\max\{0, \hat{\sigma}_{yy|x}\}$ is not an unbiased estimator, and it is difficult to formulate the truncated distribution of $\hat{\sigma}_{yy|x}$. Thus, it would also be future work to develop a more efficient estimator of the causal effect on the variance based on the small sample size. Furthermore, the assumption of Gaussian random variables may be strong. To derive the exact variance formula of the estimated causal effect on the variance under the non-Gaussian random random variables, our idea is to assume the probability distribution whose exact moments of reciprocal random variables can be derived as the explicit expressions, which is also future work. Finally, it would also be necessary to discuss the extension of our result to non-parametric SEMs in the future.

6. Competing interests

The authors declare no conflicts of interest associated with this manuscript.

Acknowledgement

We would like to acknowledge the helpful comments of the associate editor and the two anonymous reviewers. This research was partially supported by Mitsubishi Electric Corporation and Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science, Grant Number 19K11856.

Table 3. Causal 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

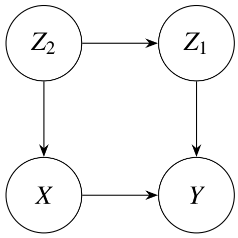

Figure 3. Causal path diagram

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

$$\mathrm{var}(\hat{\sigma}_{yy|x}) = \mathrm{var}(E(\hat{\sigma}_{yy|x}|D_x, D_z)) + E(\mathrm{var}(\hat{\sigma}_{yy|x}|D_x, D_z)), \tag{33}$$

where $E(\cdot|D_x, D_z)$ and $\mathrm{var}(\cdot|D_x, D_z)$ indicates 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 of do($X = x$) on the variance $\hat{\sigma}_{yy|x}$ of Y, 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 (33) 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 (33), note that we derive

$$\begin{aligned} E(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z) = E(\mathrm{tr}(\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}\hat{B}_{yz.xz})|D_x, D_z) &= \mathrm{tr}\left(\hat{\Sigma}_{zz}(\sigma_{yy.xz}S^{-1}_{zz.x} + B'_{yz.xz}B_{yz.xz})\right) \\ &= \sigma_{yy.xz}\mathrm{tr}(\hat{\Sigma}_{zz}S^{-1}_{zz.x}) + B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz} \end{aligned} \tag{34}$$

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

$$\begin{aligned} \mathrm{var}(\hat{B}_{yz.xz}|D_x, D_z) &= E(\hat{B}'_{yz.xz}\hat{B}_{yz.xz}|D_x, D_z) - E(\hat{B}'_{yz.xz}|D_x, D_z)E(\hat{B}_{yz.xz}|D_x, D_z) = E(\hat{B}'_{yz.xz}\hat{B}_{yz.xz}|D_x, D_z) - B'_{yz.xz}B_{yz.xz} \\ &= \sigma_{yy.xz}S^{-1}_{zz.x}, \end{aligned} \tag{35}$$

where tr(A), which is the trace of a square matrix A, represents the total sum of elements on the main diagonal of the square matrix A. Thus, noting that equation (14), $\hat{\sigma}_{yy.xz}$, is the unbiased estimator of $\sigma_{yy.xz}$, we have

$$\begin{aligned} E(\hat{\sigma}_{yy|x}|D_x, D_z) &= E\left(\hat{\sigma}_{yy.xz}\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right)|D_x, D_z\right) + E(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z) \\ &= \sigma_{yy.xz}\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right) + \mathrm{tr}(\hat{\Sigma}_{zz}S^{-1}_{zz.x})\right) + B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}. \end{aligned} \tag{36}$$

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

$$S_{zz.x}^{-1} = \left(S_{zz} - \frac{S_{zx}S_{xz}}{s_{xx}}\right)^{-1} = S_{zz}^{-1} + \frac{S_{zz}^{-1}S_{zx}S_{xz}S_{zz}^{-1}}{s_{xx.z}}. \tag{37}$$

Thus, from equation (14), noting that $\hat{\Sigma}_{zz}$ is the unbiased estimator of Σ_{zz}, we derive

$$\begin{aligned}\mathrm{tr}(\hat{\Sigma}_{zz}S_{zz.x}^{-1}) &= \frac{1}{n-1}\mathrm{tr}(S_{zz}S_{zz.x}^{-1}) = \frac{1}{n-1}\mathrm{tr}\left(I_{q,q} + \frac{S_{zx}S_{xz}S_{zz}^{-1}}{s_{xx.z}}\right) = \frac{1}{n-1}\left(q + \frac{S_{xz}S_{zz}^{-1}S_{zx}}{s_{xx.z}}\right)\\ &= \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right),\end{aligned} \tag{38}$$

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

$$E(\hat{\sigma}_{yy|x}|D_x, D_z) = \sigma_{yy.xz} + B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz} \tag{39}$$

from equation (34) together with equation (38), we derive

$$E(\hat{\sigma}_{yy|x}) = E(E(\hat{\sigma}_{yy|x}|D_x, D_z)) = \sigma_{yy.xz} + B_{yz.xz}\Sigma_{zz}B'_{yz.xz} \tag{40}$$

and

$$\mathrm{var}(E(\hat{\sigma}_{yy|x}|D_x, D_z)) = \mathrm{var}(B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}). \tag{41}$$

Equation (40) shows that $\hat{\sigma}_{yy|x}$ is the unbiased estimator of the causal effect of $\mathrm{do}(X = x)$ on the variance of Y.

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

$$\frac{(n-1)B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}}{B_{yz.xz}\Sigma_{zz}B'_{yz.xz}} \tag{42}$$

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

$$\mathrm{var}\left(\frac{(n-1)B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}}{B_{yz.xz}\Sigma_{zz}B'_{yz.xz}}\right) = 2(n-1), \tag{43}$$

i.e., we have

$$\mathrm{var}(E(\hat{\sigma}_{yy|x}|D_x, D_z)) = \mathrm{var}\left(B_{yz.xz}\hat{\Sigma}_{zz}B'_{yz.xz}\right) = \frac{2(B_{yz.xz}\Sigma_{zz}B'_{yz.xz})^2}{n-1}. \tag{44}$$

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

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

$$\frac{(n-q-2)\hat{\sigma}_{yy.xz}}{\sigma_{yy.xz}} \tag{45}$$

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

$$\begin{aligned}&\mathrm{var}(\hat{\sigma}_{yy|x}|D_x, D_z) = \mathrm{var}(\hat{\sigma}_{yy.xz}|D_x, D_z)\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right)^2 + \mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z)\\ &= \frac{2\sigma^2{}_{yy.xz}}{n-q-2}\left(1 - \frac{1}{n-1}\left(q + \frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)\right)^2 + \mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z)\\ &= \frac{2\sigma^2{}_{yy.xz}}{n-q-2}\left(\left(1 - \frac{q}{n-1}\right)^2 - 2\left(1 - \frac{q}{n-1}\right)\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{(n-1)s_{xx.z}} + \left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{(n-1)s_{xx.z}}\right)^2\right) + \mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z).\end{aligned} \tag{46}$$

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 (46), 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

$$E\left(\frac{1}{s_{xx.z}}|D_z\right) = \frac{1}{(n-q-3)\sigma_{xx.z}}. \tag{47}$$

Thus, we have

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

from

$$\mathrm{var}(\hat{B}_{xz.z}) = E(\hat{B}'_{xz.z}\hat{B}_{xz.z}) - B'_{xz.z}B_{xz.z} = \sigma_{xx.z}S_{zz}^{-1}. \tag{49}$$

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

$$E\left(\frac{1}{s_{xx.z}^2}\right) = \frac{1}{(n-q-3)(n-q-5)\sigma_{xx.z}^2}, \tag{50}$$

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

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

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

$$\begin{aligned} E\left(\mathrm{var}\left(\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}|D_z\right)\right) &= 2\sigma_{xx.z}^2E(\mathrm{tr}(S_{zz}S_{zz}^{-1}S_{zz}S_{zz}^{-1})) + 4E(\sigma_{xx.z}B_{xz.z}S_{zz}S_{zz}^{-1}S_{zz}B'_{xz.z}) \\ &= 2q\sigma_{xx.z}^2 + 4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}. \end{aligned} \tag{52}$$

Again, from

$$\mathrm{var}(B_{xz.z}S_{zz}B'_{xz.z}) = 2(n-1)(B_{xz.z}\Sigma_{zz}B'_{xz.z})^2 \tag{53}$$

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

$$\begin{aligned} &E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right) = E\left(E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2|D_z\right)\right) \\ &= \frac{2q\sigma_{xx.z}^2 + 4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} + \frac{E((q\sigma_{xx.z} + B_{xz.z}S_{zz}B'_{xz.z})^2)}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} \\ &= \frac{2q\sigma_{xx.z}^2 + 4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} + \frac{\mathrm{var}(B_{xz.z}S_{zz}B'_{xz.z})}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} + \frac{E(q\sigma_{xx.z} + B_{xz.z}S_{zz}B'_{xz.z})^2}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} \\ &= \frac{2q\sigma_{xx.z}^2 + 4(n-1)\sigma_{xx.z}B_{xz.z}\Sigma_{zz}B'_{xz.z}}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} + \frac{2(n-1)(B_{xz.z}\Sigma_{zz}B'_{xz.z})^2}{(n-q-3)(n-q-5)\sigma_{xx.z}^2} + \frac{(q\sigma_{xx.z} + (n-1)B_{xz.z}\Sigma_{zz}B'_{xz.z})^2}{(n-q-3)(n-q-5)\sigma_{xx.z}^2}. \end{aligned} \tag{54}$$

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 (46), from Mathai and Provost (1992, p.53), we have

$$\mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z) \quad = \quad \frac{2\sigma^2_{yy.xz}}{(n-1)^2}\mathrm{tr}(S_{zz}S^{-1}_{zz.x}S_{zz}S^{-1}_{zz.x}) + \frac{4\sigma_{yy.xz}}{(n-1)^2}B_{yz.xz}S_{zz}S^{-1}_{zz.x}S_{zz}B'_{yz.xz}. \tag{55}$$

From equation (37) and $\hat{B}_{xz.z} = S_{xz}S^{-1}_{zz}$, we have

$$\begin{aligned} E(\mathrm{var}(\hat{B}_{yz.xz}\hat{\Sigma}_{zz}\hat{B}'_{yz.xz}|D_x, D_z)) &= E\left(\frac{2\sigma^2_{yy.xz}}{(n-1)^2}\mathrm{tr}\left(\left(S_{zz}+\frac{S_{zx}S_{xz}}{s_{xx.z}}\right)S^{-1}_{zz.x}\right)\right) + \frac{4\sigma_{yy.xz}}{(n-1)^2}B_{yz.xz}E\left(S_{zz}+\frac{S_{zx}S_{xz}}{s_{xx.z}}\right)B'_{yz.xz} \\ &= \frac{2\sigma^2_{yy.xz}}{(n-1)^2}\left(q + 2E\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right) + E\left(\left(\frac{\hat{B}_{xz.z}S_{zz}\hat{B}'_{xz.z}}{s_{xx.z}}\right)^2\right)\right) \\ &\quad + \frac{4\sigma_{yy.xz}}{(n-1)^2}\left((n-1)B_{yz.xz}\Sigma_{zz}B'_{yz.xz} + E\left(\frac{(B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2}{s_{xx.z}}\right)\right) \end{aligned} \tag{56}$$

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

$$\begin{aligned} &E\left(E\left((B_{yz.xz}S_{zz}\hat{B}'_{xz.z})^2|D_z\right)\right) = E\left(\mathrm{var}\left(B_{yz.xz}S_{zz}\hat{B}'_{xz.z}|D_z\right) + E\left((B_{yz.xz}S_{zz}\hat{B}'_{xz.z})|D_z\right)^2\right) \\ &= \sigma_{xx.z}B_{yz.xz}E(S_{zz})B'_{yz.xz} + E((B_{yz.xz}S_{zz}B'_{xz.z})^2) = (n-1)\sigma_{xx.z}B_{yz.xz}\Sigma_{zz}B'_{yz.xz} + E((B_{yz.xz}S_{zz}B'_{xz.z})^2) \end{aligned} \tag{57}$$

Finally, from Seber (2008,p.467), we have

$$\begin{aligned} E((B_{yz.xz}S_{zz}B'_{xz.z})^2) &= B_{yz.xz}E(S_{zz}B'_{xz.z}B_{xz.z}S_{zz})B'_{yz.xz} \\ &= ((n-1)+(n-1)^2)(B_{yz.xz}\Sigma_{zz}B'_{xz.z})^2 + (n-1)(B_{xz.z}\Sigma_{zz}B'_{xz.z})(B_{yz.xz}B'_{xz.z})^2. \end{aligned} \tag{58}$$

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

References

[1] Baron, R. and Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, **51**, 1173–1182.

[2] Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology*, **17**, 37–69.

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

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

[5] 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.

[6] Chen, B. R. (2017). Graphical methods for linear structural equation modeling. Ph.D. Thesis, Department of Computer Science Department, UCLA.

[7] Cox, D. R. and Wermuth, N. (2014). *Multivariate Dependencies: Models, Analysis and Interpretation*, Chapman and Hall/CRC.

[8] Duncan, O. (1975). *Introduction to Structural Equation Models*. Academic Press, New York.

[9] 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.

[10] 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.

[11] Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*, Chapman & Hall/CRC.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.
[17] 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.
[18] 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.
[19] Kuroki, M. and Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika*, **101**, 423-437.
[20] Maathuis, M. H. and Colombo, D. (2015). A generalized back-door criterion. *Annals of Statistics*, **43**, 1060-1088.
[21] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press.
[22] Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables: Theory and Applications*, CRC Press.
[23] 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.
[24] Okuno, T., Katayama, Z., Kamigori, N., Itoh, T., Irikura, N., and Fujiwara, N. (1986). *Multivariate Data Analysis in Industry*, JUSE Press.
[25] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference, the 2nd Edition*, Cambridge University Press.
[26] Perković, E. (2018). Graphical characterizations of adjustment sets. Ph.D. Thesis, Department of Mathematics, ETH Zurich.
[27] Seber, G. A. (2008). *A Matrix Handbook for Statisticians*, John Wiley & Sons.
[28] 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.
[29] 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.
[30] 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.
[31] Tian, J. (2004). Identifying linear causal effects. *Proceeding of the 19th National Conference on Artificial Intelligence*, 104-111.
[32] Weiss, N. A., Holmes, P. T. and Hardy, M. (2006). *A Course in Probability*. Pearson Addison Wesley.
[33] Wright, S. (1923). The theory of path coefficients: A reply to Niles' criticism. *Genetics*, **8**, 239-255.
[34] Wright, S.(1934). The method of path coefficients. *Annals of Mathematical Statistics*, **5**, 161-215.

Table 4. Numerical Experiments -Normal Distribution-.

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)/(22)	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)/(22)	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

Sample size	500						1000					
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	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510
Equation (17)/(22)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
var	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.075	0.075	0.075	0.075	0.075	0.075	0.052	0.052	0.052	0.052	0.052	0.052
Kurtosis	3.041	3.041	3.037	3.037	3.041	3.041	3.030	3.030	3.030	3.030	3.030	3.030
Minimum	0.392	0.392	0.393	0.393	0.392	0.392	0.432	0.432	0.432	0.432	0.432	0.432
1st Quartile	0.491	0.491	0.491	0.492	0.491	0.491	0.497	0.497	0.497	0.497	0.497	0.497
Median	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510
3rd Quartile	0.528	0.528	0.528	0.529	0.528	0.528	0.523	0.523	0.523	0.523	0.523	0.523
Maximum	0.651	0.651	0.652	0.652	0.651	0.651	0.593	0.593	0.594	0.594	0.593	0.593

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 4. Numerical Experiments -Normal Distribution-.

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)/(22)	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)/(22)	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

Sample size	500						1000					
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	0.511	0.510	0.512	0.510	0.513	0.510	0.511	0.510	0.511	0.510	0.512
Equation (17)/(22)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
var	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Skewness	0.028	0.028	0.029	0.029	0.029	0.029	0.024	0.024	0.025	0.025	0.025	0.025
Kurtosis	2.981	2.981	2.981	2.980	2.980	2.980	3.001	3.001	2.999	2.999	3.000	2.999
Minimum	0.427	0.428	0.427	0.428	0.427	0.429	0.455	0.455	0.455	0.455	0.455	0.456
1st Quartile	0.496	0.497	0.496	0.498	0.496	0.499	0.500	0.501	0.500	0.501	0.500	0.502
Median	0.510	0.511	0.510	0.512	0.510	0.513	0.510	0.511	0.510	0.511	0.510	0.511
3rd Quartile	0.524	0.525	0.524	0.526	0.524	0.527	0.520	0.521	0.520	0.521	0.520	0.521
Maximum	0.601	0.603	0.600	0.603	0.600	0.604	0.577	0.578	0.577	0.578	0.577	0.578

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 4. Numerical Experiments -Normal Distribution-.

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)/(22)	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)/(22)	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

Sample size	500						1000					
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	1.119	1.118	1.121	1.117	1.121	1.118	1.118	1.118	1.119	1.118	1.120
Equation (17)/(22)	0.006	0.006	0.007	0.007	0.006	0.006	0.003	0.003	0.003	0.003	0.003	0.003
var	0.005	0.005	0.005	0.005	0.005	0.005	0.002	0.002	0.003	0.003	0.002	0.002
Skewness	0.100	0.100	0.155	0.154	0.129	0.128	0.097	0.097	0.117	0.117	0.111	0.111
Kurtosis	2.994	2.994	3.001	3.001	2.993	2.993	3.026	3.026	3.053	3.053	3.040	3.040
Minimum	0.860	0.861	0.844	0.847	0.848	0.852	0.932	0.933	0.909	0.911	0.922	0.923
1st Quartile	1.071	1.073	1.066	1.070	1.069	1.073	1.085	1.086	1.082	1.084	1.084	1.085
Median	1.116	1.118	1.116	1.119	1.116	1.120	1.117	1.118	1.116	1.118	1.117	1.119
3rd Quartile	1.163	1.164	1.166	1.170	1.164	1.168	1.149	1.150	1.152	1.154	1.151	1.153
Maximum	1.416	1.418	1.482	1.486	1.434	1.438	1.316	1.316	1.348	1.350	1.342	1.344

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 4. Numerical Experiments -Normal Distribution-.

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)/(22)	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)/(22)	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

Sample size	500						1000					
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	1.001	1.000	1.003	1.000	1.003	1.000	1.001	1.000	1.002	1.000	1.002
Equation (17)/(22)	0.005	0.005	0.005	0.005	0.005	0.005	0.002	0.002	0.003	0.003	0.002	0.002
var	0.004	0.004	0.004	0.004	0.004	0.004	0.002	0.002	0.002	0.002	0.002	0.002
Skewness	0.125	0.125	0.173	0.173	0.142	0.141	0.089	0.089	0.123	0.123	0.107	0.107
Kurtosis	3.052	3.052	3.063	3.063	3.050	3.050	2.982	2.982	3.023	3.023	3.005	3.005
Minimum	0.773	0.774	0.752	0.754	0.768	0.770	0.825	0.825	0.797	0.799	0.815	0.816
1st Quartile	0.958	0.960	0.954	0.957	0.956	0.959	0.970	0.971	0.967	0.969	0.969	0.970
Median	0.999	1.000	0.998	1.001	0.998	1.001	0.999	1.000	0.999	1.000	0.999	1.001
3rd Quartile	1.040	1.041	1.043	1.046	1.042	1.045	1.029	1.030	1.031	1.033	1.030	1.031
Maximum	1.285	1.287	1.299	1.303	1.291	1.294	1.191	1.191	1.235	1.236	1.224	1.225

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 4. Numerical Experiments -Normal Distribution-.

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)/(22)	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)/(22)	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

Sample size	500						1000					
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	6.914	6.890	6.952	6.893	6.943	6.889	6.902	6.892	6.923	6.892	6.917
Equation (17)/(22)	0.807	0.798	1.736	1.710	1.415	1.395	0.401	0.399	0.861	0.855	0.702	0.697
var	0.710	0.710	1.618	1.619	1.309	1.310	0.352	0.352	0.815	0.815	0.654	0.655
Skewness	0.227	0.227	0.301	0.301	0.281	0.281	0.174	0.174	0.245	0.245	0.214	0.214
Kurtosis	3.106	3.106	3.148	3.148	3.114	3.114	3.041	3.041	3.089	3.089	3.059	3.059
Minimum	3.987	4.011	2.650	2.716	2.966	3.016	4.867	4.880	3.770	3.796	4.135	4.160
1st Quartile	6.311	6.336	5.997	6.058	6.096	6.146	6.482	6.494	6.266	6.297	6.332	6.357
Median	6.855	6.879	6.828	6.891	6.837	6.887	6.871	6.883	6.857	6.888	6.866	6.891
3rd Quartile	7.437	7.462	7.710	7.772	7.628	7.679	7.282	7.294	7.480	7.512	7.421	7.447
Maximum	10.770	10.797	13.431	13.498	13.226	13.273	9.556	9.569	10.954	10.983	10.449	10.476

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 5. Numerical Experiments -Uniform Distribution-.

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	0.510	0.513	0.510	0.545	0.510	0.515	0.510	0.511	0.510	0.521	0.510	0.511
Equation (17)/(22)	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	0.044	0.044	0.051	0.054	0.044	0.044	0.016	0.016	0.017	0.017	0.016	0.016
Skewness	0.596	0.596	0.762	0.722	0.597	0.596	0.328	0.328	0.337	0.324	0.328	0.328
Kurtosis	3.376	3.375	4.638	4.628	3.368	3.366	3.111	3.111	3.116	3.105	3.113	3.113
Minimum	0.033	0.034	-0.066	0.030	0.031	0.036	0.098	0.098	0.089	0.093	0.097	0.099
1st Quartile	0.357	0.360	0.346	0.377	0.357	0.362	0.421	0.422	0.419	0.430	0.421	0.423
Median	0.488	0.490	0.485	0.521	0.488	0.493	0.503	0.504	0.502	0.514	0.503	0.505
3rd Quartile	0.641	0.643	0.646	0.687	0.640	0.646	0.591	0.592	0.593	0.605	0.591	0.593
Maximum	1.613	1.614	3.778	3.951	1.617	1.620	1.205	1.206	1.177	1.192	1.214	1.217

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	0.511	0.510	0.516	0.510	0.511	0.510	0.510	0.510	0.512	0.510	0.510
Equation (17)/(22)	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	0.008	0.008	0.008	0.008	0.008	0.008	0.004	0.004	0.004	0.004	0.004	0.004
Skewness	0.231	0.230	0.235	0.230	0.230	0.230	0.170	0.170	0.173	0.172	0.170	0.170
Kurtosis	3.061	3.061	3.071	3.068	3.061	3.061	3.006	3.006	3.010	3.009	3.006	3.006
Minimum	0.196	0.196	0.187	0.189	0.195	0.196	0.257	0.258	0.249	0.251	0.258	0.258
1st Quartile	0.450	0.450	0.449	0.455	0.450	0.451	0.467	0.467	0.467	0.469	0.467	0.467
Median	0.506	0.507	0.506	0.512	0.506	0.507	0.508	0.508	0.508	0.510	0.508	0.508
3rd Quartile	0.568	0.568	0.568	0.574	0.568	0.568	0.551	0.551	0.551	0.554	0.551	0.551
Maximum	0.932	0.933	0.934	0.941	0.932	0.933	0.798	0.798	0.793	0.796	0.798	0.798

Sample size	500						1000					
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	0.510	0.510	0.511	0.510	0.510	0.510	0.510	0.510	0.510	0.510	0.510
Equation (17)/(22)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
var	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Skewness	0.140	0.140	0.141	0.141	0.141	0.141	0.087	0.087	0.088	0.088	0.087	0.087
Kurtosis	3.025	3.025	3.028	3.028	3.025	3.025	3.063	3.063	3.063	3.063	3.063	3.063
Minimum	0.389	0.389	0.389	0.389	0.389	0.389	0.412	0.412	0.412	0.412	0.412	0.412
1st Quartile	0.488	0.488	0.488	0.488	0.488	0.488	0.494	0.494	0.494	0.495	0.494	0.494
Median	0.509	0.509	0.509	0.510	0.509	0.509	0.509	0.509	0.509	0.510	0.509	0.510
3rd Quartile	0.532	0.532	0.532	0.532	0.532	0.532	0.525	0.525	0.525	0.525	0.525	0.525
Maximum	0.664	0.664	0.665	0.666	0.664	0.664	0.616	0.616	0.616	0.616	0.616	0.616

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 5. Numerical Experiments -Uniform Distribution-.

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	0.509	0.599	0.509	0.647	0.508	0.731	0.510	0.540	0.510	0.555	0.510	0.578
Equation (17)/(22)	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	0.038	0.051	0.058	0.078	0.083	0.116	0.010	0.011	0.011	0.013	0.012	0.014
Skewness	0.714	0.813	2.428	2.593	3.456	3.531	0.168	0.165	0.386	0.361	0.445	0.379
Kurtosis	6.253	6.574	41.824	39.565	90.701	58.000	3.052	3.044	3.768	3.702	3.988	3.767
Minimum	-1.852	0.003	-3.606	0.003	-7.368	0.004	0.153	0.162	0.154	0.169	0.153	0.173
1st Quartile	0.371	0.439	0.354	0.462	0.341	0.514	0.440	0.466	0.437	0.476	0.437	0.496
Median	0.494	0.582	0.482	0.618	0.474	0.690	0.507	0.537	0.506	0.550	0.505	0.574
3rd Quartile	0.629	0.739	0.628	0.788	0.629	0.887	0.576	0.610	0.577	0.627	0.577	0.654
Maximum	2.946	3.659	8.443	9.805	11.662	12.334	1.056	1.096	1.396	1.458	1.366	1.476

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	0.526
Equation (17)/(22)	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	0.005	0.005	0.005	0.005	0.005	0.005	0.002	0.002	0.002	0.002	0.002	0.002
Skewness	0.106	0.106	0.168	0.163	0.177	0.165	0.068	0.068	0.084	0.083	0.085	0.083
Kurtosis	3.006	3.005	3.117	3.107	3.129	3.105	3.018	3.018	3.024	3.025	3.024	3.025
Minimum	0.252	0.262	0.257	0.268	0.255	0.273	0.314	0.318	0.317	0.323	0.317	0.326
1st Quartile	0.464	0.476	0.462	0.482	0.462	0.492	0.478	0.485	0.478	0.488	0.478	0.493
Median	0.509	0.523	0.509	0.530	0.508	0.540	0.509	0.516	0.509	0.520	0.509	0.525
3rd Quartile	0.555	0.571	0.555	0.578	0.556	0.590	0.541	0.549	0.542	0.552	0.542	0.558
Maximum	0.825	0.849	0.942	0.969	0.942	0.981	0.709	0.720	0.757	0.776	0.765	0.791

Sample size	500						1000					
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	0.511	0.510	0.512	0.510	0.513	0.510	0.511	0.510	0.511	0.510	0.511
Equation (17)/(22)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
var	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Skewness	0.139	0.139	0.140	0.140	0.140	0.140	0.086	0.085	0.086	0.086	0.086	0.086
Kurtosis	3.034	3.033	3.036	3.036	3.036	3.036	3.025	3.025	3.024	3.024	3.024	3.024
Minimum	0.373	0.374	0.373	0.374	0.373	0.375	0.420	0.420	0.419	0.420	0.419	0.421
1st Quartile	0.488	0.489	0.488	0.490	0.488	0.491	0.494	0.495	0.494	0.495	0.494	0.496
Median	0.509	0.511	0.509	0.511	0.509	0.512	0.510	0.510	0.510	0.511	0.510	0.511
3rd Quartile	0.531	0.533	0.531	0.533	0.531	0.535	0.525	0.526	0.525	0.526	0.525	0.527
Maximum	0.643	0.645	0.645	0.647	0.644	0.648	0.613	0.613	0.612	0.614	0.612	0.614

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 5. Numerical Experiments -Uniform Distribution-.

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	1.115	1.225	1.114	1.352	1.115	1.389	1.117	1.153	1.118	1.194	1.117	1.200
Equation (17)/(22)	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	0.328	0.354	0.468	0.559	0.456	0.518	0.102	0.105	0.126	0.135	0.115	0.120
Skewness	1.505	1.502	2.207	2.202	2.133	2.265	0.570	0.557	0.850	0.815	0.724	0.687
Kurtosis	14.641	15.279	16.858	16.258	19.287	21.102	3.562	3.549	4.391	4.313	4.023	3.963
Minimum	-0.946	0.046	-6.136	0.047	-3.885	0.050	0.179	0.188	0.177	0.190	0.177	0.197
1st Quartile	0.702	0.799	0.658	0.849	0.660	0.905	0.889	0.923	0.866	0.934	0.876	0.956
Median	1.025	1.138	0.978	1.216	0.994	1.272	1.089	1.125	1.071	1.149	1.080	1.165
3rd Quartile	1.425	1.549	1.409	1.689	1.428	1.725	1.312	1.351	1.321	1.405	1.316	1.406
Maximum	15.633	16.584	13.776	14.530	16.626	18.175	3.128	3.184	3.789	4.023	3.433	3.531

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	1.119	1.136	1.119	1.155	1.119	1.158	1.117	1.126	1.117	1.135	1.117	1.136
Equation (17)/(22)	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	0.048	0.049	0.058	0.060	0.053	0.054	0.023	0.023	0.028	0.028	0.025	0.026
Skewness	0.396	0.392	0.580	0.566	0.500	0.486	0.258	0.256	0.375	0.370	0.323	0.319
Kurtosis	3.257	3.254	3.624	3.606	3.473	3.462	3.100	3.099	3.233	3.227	3.182	3.179
Minimum	0.385	0.396	0.374	0.391	0.386	0.409	0.601	0.609	0.569	0.581	0.581	0.596
1st Quartile	0.965	0.981	0.949	0.982	0.956	0.995	1.011	1.019	1.000	1.017	1.006	1.024
Median	1.105	1.122	1.097	1.133	1.100	1.140	1.111	1.119	1.107	1.125	1.109	1.128
3rd Quartile	1.257	1.275	1.265	1.304	1.262	1.303	1.216	1.225	1.222	1.241	1.220	1.239
Maximum	2.289	2.307	2.573	2.628	2.436	2.477	1.930	1.941	1.912	1.935	1.954	1.977

Sample size	500						1000					
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	1.118	1.120	1.118	1.121	1.118	1.122	1.118	1.119	1.118	1.120	1.118	1.120
Equation (17)/(22)	0.006	0.006	0.007	0.007	0.006	0.006	0.003	0.003	0.003	0.003	0.003	0.003
var	0.006	0.006	0.007	0.007	0.006	0.006	0.003	0.003	0.003	0.003	0.003	0.003
Skewness	0.162	0.162	0.209	0.209	0.185	0.184	0.108	0.108	0.134	0.134	0.117	0.117
Kurtosis	3.013	3.013	3.077	3.077	3.041	3.041	3.048	3.048	3.062	3.061	3.052	3.052
Minimum	0.842	0.844	0.832	0.835	0.831	0.834	0.908	0.908	0.889	0.891	0.909	0.911
1st Quartile	1.065	1.067	1.062	1.065	1.063	1.067	1.081	1.082	1.079	1.081	1.080	1.082
Median	1.116	1.118	1.115	1.119	1.115	1.119	1.117	1.118	1.117	1.119	1.117	1.119
3rd Quartile	1.169	1.170	1.172	1.175	1.170	1.174	1.153	1.154	1.156	1.158	1.154	1.156
Maximum	1.513	1.514	1.533	1.537	1.542	1.546	1.378	1.379	1.366	1.368	1.370	1.372

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 5. Numerical Experiments -Uniform Distribution-.

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	1.091	1.000	1.206	1.000	1.223	0.999	1.029	0.998	1.064	0.999	1.067
Equation (17)/(22)	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	0.264	0.282	0.379	0.452	0.362	0.405	0.083	0.085	0.104	0.111	0.093	0.097
Skewness	1.174	1.164	2.124	2.164	2.103	2.265	0.590	0.580	0.899	0.865	0.745	0.712
Kurtosis	5.852	5.955	15.041	16.253	17.794	20.338	3.690	3.674	4.593	4.522	4.306	4.241
Minimum	-0.584	0.047	-3.721	0.046	-2.829	0.054	0.180	0.190	0.180	0.190	0.180	0.199
1st Quartile	0.631	0.709	0.585	0.750	0.594	0.794	0.795	0.823	0.768	0.827	0.781	0.847
Median	0.920	1.011	0.872	1.081	0.890	1.118	0.973	1.003	0.956	1.022	0.967	1.036
3rd Quartile	1.278	1.378	1.271	1.515	1.278	1.521	1.176	1.207	1.178	1.252	1.177	1.250
Maximum	5.910	6.216	10.743	13.037	13.002	14.269	2.977	3.034	3.134	3.304	4.450	4.548

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	0.999	1.013	0.999	1.030	0.999	1.030	1.000	1.007	1.000	1.016	1.000	1.016
Equation (17)/(22)	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	0.039	0.040	0.048	0.049	0.043	0.044	0.019	0.019	0.023	0.024	0.021	0.021
Skewness	0.410	0.406	0.575	0.563	0.489	0.478	0.269	0.267	0.383	0.379	0.323	0.319
Kurtosis	3.272	3.271	3.546	3.528	3.433	3.424	3.136	3.136	3.260	3.256	3.195	3.193
Minimum	0.411	0.419	0.398	0.417	0.399	0.423	0.534	0.540	0.506	0.517	0.516	0.529
1st Quartile	0.859	0.872	0.844	0.873	0.853	0.884	0.904	0.911	0.893	0.908	0.899	0.914
Median	0.986	1.000	0.978	1.009	0.981	1.014	0.994	1.001	0.991	1.006	0.993	1.008
3rd Quartile	1.125	1.139	1.133	1.167	1.127	1.160	1.090	1.097	1.097	1.113	1.093	1.108
Maximum	2.107	2.124	2.383	2.423	2.206	2.258	1.660	1.669	1.796	1.815	1.792	1.809

Sample size	500						1000					
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	1.001	1.000	1.003	1.000	1.003	1.000	1.001	1.000	1.001	1.000	1.001
Equation (17)/(22)	0.005	0.005	0.005	0.005	0.005	0.005	0.002	0.002	0.003	0.003	0.002	0.002
var	0.005	0.005	0.005	0.005	0.005	0.005	0.002	0.002	0.003	0.003	0.002	0.002
Skewness	0.167	0.166	0.208	0.207	0.182	0.182	0.117	0.117	0.141	0.141	0.132	0.132
Kurtosis	3.075	3.075	3.094	3.094	3.079	3.079	3.071	3.071	3.063	3.063	3.072	3.072
Minimum	0.759	0.760	0.732	0.734	0.745	0.748	0.826	0.826	0.809	0.810	0.810	0.811
1st Quartile	0.953	0.954	0.949	0.952	0.952	0.955	0.967	0.968	0.964	0.966	0.966	0.967
Median	0.998	1.000	0.997	1.000	0.998	1.001	0.999	1.000	0.999	1.000	0.999	1.000
3rd Quartile	1.045	1.046	1.048	1.051	1.046	1.049	1.032	1.033	1.034	1.036	1.033	1.034
Maximum	1.318	1.319	1.351	1.355	1.321	1.324	1.214	1.215	1.243	1.245	1.234	1.236

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.

Table 5. Numerical Experiments -Uniform Distribution-.

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	6.886	8.826	6.914	11.437	6.948	11.304	6.887	7.460	6.871	8.269	6.880	8.064
Equation (17)/(22)	89.284	39.921	273.079	85.493	255.709	69.743	20.461	15.968	48.386	34.197	39.013	27.897
var	73.274	76.123	189.405	194.614	188.836	194.565	17.568	17.847	41.738	42.406	34.538	35.049
Skewness	2.759	3.012	2.518	3.008	3.165	3.946	1.092	1.105	1.493	1.515	1.374	1.389
Kurtosis	23.382	25.464	16.877	18.783	31.337	46.078	4.870	4.905	6.484	6.576	5.769	5.820
Minimum	-29.946	0.020	-204.962	0.044	-135.808	0.085	-1.136	0.232	-3.933	0.172	-3.418	0.181
1st Quartile	1.244	2.876	-1.183	2.255	-0.891	2.354	3.833	4.380	2.098	3.428	2.524	3.670
Median	4.734	6.445	2.855	6.685	3.075	6.710	6.159	6.723	5.340	6.715	5.569	6.724
3rd Quartile	10.096	11.981	11.158	15.365	11.081	15.100	9.150	9.727	9.972	11.356	9.786	11.001
Maximum	161.940	167.812	179.531	203.065	370.455	460.304	39.672	40.183	78.119	80.845	53.620	56.525

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	6.893	7.158	6.908	7.563	6.912	7.453	6.900	7.028	6.896	7.214	6.896	7.155
Equation (17)/(22)	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	7.738	7.796	18.267	18.403	14.919	15.022	3.722	3.737	8.604	8.641	6.948	6.975
Skewness	0.737	0.739	1.012	1.016	0.936	0.937	0.512	0.513	0.738	0.739	0.664	0.664
Kurtosis	3.858	3.859	4.484	4.491	4.262	4.260	3.383	3.383	3.820	3.821	3.638	3.638
Minimum	0.201	0.411	-0.649	0.356	-0.319	0.466	1.036	1.161	0.209	0.505	0.285	0.491
1st Quartile	4.886	5.147	3.765	4.407	4.079	4.608	5.525	5.652	4.789	5.100	4.999	5.256
Median	6.547	6.814	6.195	6.837	6.314	6.855	6.738	6.865	6.537	6.855	6.613	6.870
3rd Quartile	8.548	8.820	9.300	9.963	9.105	9.660	8.090	8.221	8.629	8.950	8.475	8.737
Maximum	29.321	29.573	36.865	37.688	32.626	33.335	17.138	17.308	24.775	25.115	21.457	21.745

Sample size	500						1000					
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	6.887	6.912	6.898	6.961	6.895	6.945	6.890	6.903	6.893	6.924	6.892	6.917
Equation (17)/(22)	0.807	0.798	1.736	1.710	1.415	1.395	0.401	0.399	0.861	0.855	0.702	0.697
var	0.817	0.819	1.748	1.751	1.429	1.432	0.400	0.400	0.856	0.857	0.700	0.701
Skewness	0.279	0.280	0.358	0.359	0.336	0.336	0.181	0.181	0.242	0.242	0.238	0.238
Kurtosis	3.143	3.143	3.191	3.191	3.189	3.189	3.068	3.068	3.100	3.100	3.078	3.078
Minimum	3.589	3.610	2.812	2.870	3.026	3.077	4.681	4.692	3.533	3.563	3.975	3.999
1st Quartile	6.264	6.289	5.969	6.032	6.054	6.103	6.452	6.464	6.253	6.285	6.310	6.335
Median	6.842	6.867	6.822	6.882	6.833	6.882	6.872	6.884	6.855	6.886	6.858	6.883
3rd Quartile	7.471	7.497	7.744	7.807	7.662	7.712	7.308	7.321	7.495	7.527	7.437	7.463
Maximum	11.430	11.455	13.691	13.760	13.156	13.217	10.122	10.135	11.468	11.504	10.926	10.956

Unbiased: unbiased estimator; Consist: consistent estimator; Estimates: the sample mean from 50000 estimated causal effects on the variance; (17)/(22): the exact and asymptotic variances derived from equations (17) and (22) with Table 3; Var: empirical variances from 50000 estimated causal effects on the variance.