Referee Report on "An Unbiased Estimator of the Causal Effect on the Variance in Gaussian Linear Structural Equation Models"

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

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

Associate Editor

> Please read the comments given by both reviewers. To summarize, the motivation of the problem is questionable, at least it has not been explained clearly through the discussed example. Moreover, the proposed method relies on strong assumptions, and its robustness has not been studied through simulations. There are also issues on the clarity of the presentation, such as unnecessary extra notations, unclear conditions for the identifiability, unclear usefulness/significance of the main results, etc. Also, the simulations are not sufficient to demonstrate the advantage of the new method over existing ones.
> The authors are recommended to give a more convincing argument on the motivation of the problem. Key assumptions such as Gaussianity, linearity, etc, should be discussed, and numerical simulations should be conducted to test the robustness of the proposed methods on these key assumptions. The presentation of the paper requires improvement. In particular, the main contributions of the paper should be explained carefully so that the readers can understand their significance and relevance. More extensive experiments need to be included to demonstrate the improvement of the proposed method over existing ones.

We have revised the manuscript in accordance with the reviewers' comments. We think that the comments of the associate editor are included in the reviewers' comments, but please let me know if the associate editor thinks our revision is unsatisfactory for the associate editor.

Reviewer1

Major Comments

> 1. The authors proposed an estimator for the potential variance var(Y |do(X = x)). I wonder why this estimator is useful in practice. The authors mention two examples to motivate the target estimand. However, I think both bias and variance matter in the motivation examples. For instance, in the hyperglycemia example, the goal is to have the blood glucose at an acceptable range, where both the mean and variance play an important role. Therefore, I suggest the authors consider a more related motivating example. In addition, it would be helpful to discuss how to incorporate both treatment effects to more reliably investigate the effects of the treatment.

According to the reviewer's comments, as a motivating example, we have described a case study of setting up

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

coating conditions for car bodies, reported by Okuno et al. (1986). Please see Section 1.2.

> 2.This paper assumes linear SEMs hold on a known DAG. I am curious about when this assumption is true in practice and how to test the validity of this assumption (e.g., linearity, Gaussian noise, known DAG). How robust is the proposed estimator if parts of the assumption fail? The simulations study the ideal case that the assumption is valid. More extensive experiments should be conducted for robustness.

We have conducted numerical experiments to focus on how robust is the proposed estimator when the assumption of the Gaussian random disturbances is violated, because other reviewer mentioned the violation of the Gaussian random disturbances and much space is required to describe all that the reviewer mentioned. If the reviewer thins that other assumption should be discussed, please let us know.

> 3. The linear SEM is a popular tool in mediation analysis. I suggest the authors clarify the connection more clearly. In addition, the proposed estimator works for estimating the total effects. Is that possible to generalize it to estimate the direct effects?

Regarding the first comments from the reviewer, we have added the following sentences in the second paragraph of Section 1: “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). "

Regarding the second comments from the reviewer, it would be possible to generalize our results to estimate the direct effects in terms of the joint intervention, which we have already stated as one of the future work in Section 5.

> 4. The definition of a collider is not 100% correct in Section 2.2. In causal graphs, a variable is a collider when it is causally influenced by two or more variables, not only by two.

According to the reviewer’s comments, we have revised the definition of a collider as follows:

‘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."

> 5. The authors assume the readers have a strong background in causal graphs. To attract a broader audience, I suggest the authors include some examples of DAGs when introducing the definitions in Section 2.3.

According to the reviewer’s comments, we have included some examples of DAGs after Definitions 1, 2 and 3.

> 6. What assumptions are needed for the identifiability of estimating the treatment effects besides the backdoor criterion?

As we stated, we need the assumption that $\sigma_{xx.z} \neq 0$ and $\Sigma_{zz.x}$ is a positive definite matrix to estimate the treatment effect.

> 7. Section 3 lists several main results of the paper. I suggest the authors also include discussions about why we need each theorem, what we can learn from them and how we can use them.

According to the reviewer’s comments, we have added following sentences in Section 5:

“In contrast, we showed in Theorem 1 that the variance estimator in equation (16) performs better than that in (21) in finite samples. Theorem 1 implies that the proposed estimator would help us avoid the problem, and the results of this paper would help statistical practitioners to predict appropriately what would happen to the outcome variable when conducting the external intervention. Furthermore, Theorem 2 shows that the asymptotic estimation accuracy of the estimated causal effect on the variance depends on the selection of covariates 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. "

> 8. In the simulation studies, the authors studied very small sample sizes n = 10, 25, 50, 100 in order to compare the new unbiased estimator with the existed asymptotic estimator. When the sample size is 100, which is too small in practice, there is no significant improvement. I suggest the authors consider at least one more realistic sample size, e.g., n = 1000.

We have added the simulation studies with the sample sizes $n = 500$ and 1000.

> 9. As the authors noticed, the estimated variance can be negative. They proposed a simple fix to take the maximum of 0 and the estimated variance for the final estimation. In this case, how can one estimate the variance for the estimated treatment effects on the variance?

According to the reviewer's comments, We have revised the last two sentences as follows:
"However, it is difficult to formulate the truncated distribution of $\hat{\sigma}_{yy|x}$. Thus, although the use of the resampling method would be better to estimate the variance of $\max\{0, \hat{\sigma}_{yy|x}\}$, $\max\{0, \hat{\sigma}_{yy|x}\}$ is not an unbiased estimator in the first place. Thus, it would also be future work to develop a more efficient estimator of the causal effect on the variance based on the finite sample size. "

Minor Comments

> 1.The authors defined two classes of notations when there is a path from a to b – parent-child and ancestor-descendant – and use them exchangeably in the paper. To avoid confusion, I suggest the authors unify the terminologies and use one of them.

It is difficult for us to unify the terminology and one of them because equation (2) is formulated based on the parents of V_i but the back-door criterion is defined based on non-"descendants" of X. That is to say, the random variables of the RHS of equation (2) should be restricted to a set of vertices that have a directed edge to V_i. Non-descendants of X in the back-door criterion should not be restricted to a set of vertices that have a directed edge to X. However, taking the reviewer's comment into account, we have added the following sentences in Section 2.1: "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. "

Please let us know if the reviewer has any suggestion.

> 2. What is m? It is the number of nodes in a directed path a → b in Section 2.2 and the number of vertices in the DAG in Section 2.3. Please change notations.

We have revised "a path between a and b" to "a path between a and b with the length m", and change the number of variables from "m" to "p".

Reviewer2

> **1. Causal interpretation.** I feel the current definition of causal effect is not as convincing as the average treatment effect (i.e. the effect defined with the difference of means.) One reason is that the current effect defined by variance cannot be written in terms of the potential outcome. Am I right?

Pearl (2010) states "P($Y = y$|do($X = x$)) is equivalent to $P(Y_x = y)$." Thus, as $E(Y|\mathrm{do}(X = x))$ is represented by $E(Y_x)$ from the viewpoint of the potential outcome approach, $\mathrm{var}(Y|\mathrm{do}(X = x))$ is represented by $\mathrm{var}(Y_x)$. Here, we have judged that it would be better not to state such equivalence between them because we need to define some notation (potential outcomes) in order to discuss it. We can do it if the reviewer thinks it is important to state that.

> **2. Motivation and real data.** This is also related to the first point. One of my main question is why the causal effect on the variance of the outcome variable is interesting. The authors briefly mentioned this in the introduction. I would suggest the authors expand that discussion as this is a bit unusual from my perspective. Also, a real data application would be nice.

According to the reviewer's comments, as a real data application, we described a case study of setting up coating conditions for car bodies, reported by Okuno et al. (1986). Please see Section 1.2.

> **3. Prior knowledge of Z.** The whole paper assumes the knowledge of Z in the back door criterion. So this is a very big assumption. They should be estimated from the real data, right? Then this estimation creates additional uncertainty?

Yes, because the "back door criterion" is based on a causal but not statistical concept. Generally, the "back door criterion" can not be tested from statistical data. Thus, we have added the following sentences after Definition 3:

As seen from the description of Definition 3, the back-door criterion is not a statistical concept, and thus the "back door criterion" can not be tested through statistical data.

> **4. Gaussian.** The Gaussianality assumption seems quite strong. Any idea about relaxing it?

Exactly to say, it is not necessary to assume the Gaussian random disturbances to derive the unbiased estimator of the causal effect on the variance: both the linearity and the existence of the moments are necessary. Althugh the paper uses the assumption of the Gaussian random disturbances to derive the exact variance of the unbiased estimator, the assumption could be relaxed to the assumption that explicit formula of the moment of the reciprocal random variables are derived.

> **5. Notations.** The notations in (6) and (11) are very hard to digist. However, they are just used to facilitate the OLS estimator, right? If so, I suggest the authors to use less self-defined notations in (6) and (17) but use more transparent notations to improve readability.

We have revised the notation. Please see Sections 2.3 and 3 if the reviewer is satisfactory.

> **6. Contributions.** From my understanding, one main contribution is that the variance estimator in equation (15) performs better than that in (20), in finite samples. Please highlight this more if that is the main contribution and explain why it is the case.

We have added the following sentence in Section 5:

"In contrast, we showed that the variance estimator in equation (16) performs better than that in (21) in finite samples."

> 7. As a minor point, the authors use $n^{-2} \approx 0$, would this be the same with $n^{-1} \approx 0$?

We have revised $n^{-2} \approx 0$, to $n^{-1} >> n^{-2} \approx 0$.

> The results in (25) and (26) are confusing: in (25), with the additional variable z_1, the variance decreases; in (26), with the additional variable z_2, the variance increases. If this is correct, I would suggest the authors providing more intutions about this.

We have added the following sentences after Lemma 1:

"Intuitively, equation (26) shows that the estimation accuracy could be improved because the residual variance decrease by adding $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_1$ is not correlated with X given $\boldsymbol{Z}_2$. Equation (27) shows that the estimation accuracy could be worse because the residual variance does not decrease by adding $\boldsymbol{Z}_1$ and the multicollinearity occurs by adding $\boldsymbol{Z}_2$."