We thanks the reviewers for the insightful comments and helpful suggestions. Please see below for our response.

**Reviewer 2**: Regarding the baseline methods in the experiments: Original LiNGAM assumes that there is no confounder. So the issue is that it is not clear how to compare its result with the groundtruth graph (with confounders). For LiNGAM with latent variables (LvLiGNAM) by Hoyer et al. (2008), the confounders are assumed to be independent, making it impossible to discovery their connectivity. Nevertheless, we include the results of LvLiNGAM for comparison (Table 1). For clustering methods, clearly, different assumptions for clustering will lead to different clustering results. K-means clusters are used to divide data points into groups, but in our case, we divide variables into groups.

**Reviewer 3**: 1. Regarding "the setting is limited": We totally agree, and at the same time would like to mention that linear latent variable models are common in the social sciences and ought to be more common in economics and elsewhere and that most, if not all, of the methods available for such problems have stronger (in one or another dimension) assumptions than the Triad method. 2. Regarding “compare with latent tree algorithm”: Thanks for raising this issue. Generally speaking, they use the covariance information to recover the structure of the variables, ignoring non-Gaussinaity. Thus, our method can recover arbitrary DAG structure of latent variables, while the tree/latent tree learning algorithm can only solve the problems with the tree structure of latent variables the structures. We will include the results of three classic latent tree methods (neighbor-joining (NJ), by Saitou and Nei ( 1987), and recursive grouping (RG) and CLGrouping (CLNJ), by Choi et al. (2011)) in Table 1. 3. Regarding “the algorithm is sound or complete”: The soundness and completeness of the algorithms were implied in the theoretical results, and will be made more explicit. In detail, Theorem 2 and Proposition 1 ensure the correctness of Phase 1 of our method (Algorithm 1), and Theorem 1 and Proposition 2 ensure the correctness of Phase 2 of our method (Algorithm 2). We will improve the presentation. 4. Regarding “find an equivalent class or the true graph”: We are able to go beyond the equivalence class because of non-Gaussian of the data. we can uniquely recover the structure, including the structure over the latent variables, under our assumptions. 5. Regarding “noise with fifth power”: Here, we set it to ensure the noise is clearly non-Gaussian. We also varied the powers and the results are included in the revised paper. Overall, we find that not surprisingly, the more non-Gaussian, the better the performance.

**Reviewer 4**: 1. Regarding "are there a class of graphs that the present work would not be able to recover but previous methods would?": There exist graphs following different assumptions from ours that can not recovered by our method. For instance, if the confounders are independent while the observe variables have directed edges in between, LvLiNGAM might be able to recover the graph, but our method cannot. However, under our model assumptions, there does not exist any graph that can be recovered by previous methods but not ours. 2. Regarding “unclear to me if it is more general than ICA-type methods”: Yes, you are total right. It’s not necessary more general than that–we allow direct causal relations between latent variables and rely on different assumptions. To the best of our knowledge, LvLiNGAM assumes that the latent variables are independent. 3. Regarding “require a single latent common cause and no observed parents”: Here, our method can find the true graph under the purity assumption. As we discussed in our paper, if this assumption is violated, our method can still find an pure structure equivalent to the underlying causal structure. 4. Regarding “clarify there is no direct causal relation between observed variables”: The latter is right, i.e., one cannot be a parent of the other. We will emphasize this point in the revision. 5. Regarding “generalizing this method further”: Thanks for the interesting example. Definition 1 does not allow directed edges between two observed variables. We agree that it is nontrivial to generalize the method further, which we have been working on. 6. Regarding “could it not be relaxed to at most one noise term is Gaussian”: Thanks for your insightful comments. Yes, this assumption can be relaxed to at most one noise term is Gaussian for observed variables, but not the latent variables. This can be seen from the proof, and will be discussed in the paper. 7. Regarding “directed connected mean existence of a directed path”: Yes, it means directed path between $L_a$ and $L_b$ (there might be some intermediate variables in between). 8. Regarding the replacement of the latent variable with an observed: Yes, it relies on the linearity assumption. Following your suggestions, we will explain why in the paper. 9. Regarding “comparison against an ICA-type method”: Thanks for the helpful suggestions. Please refer to the explanation in lines 3-7.

Table 1: Evaluation of output latent variables (due to space limitation, only results on case 1 and case 2 are reported )

| | | Latent omission | | | | Latent commission | | | | Mismeasurements | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Algorithm | | NJ | RG | CLNJ | LvLiNGAM | NJ | RG | CLNJ | LvLiNGAM | NJ | RG | CLNJ | LvLiNGAM |
| *Case 1* | 500 | 0.40(3) | 0.45(4) | 0.45(4) | - | 0.00(3) | 0.00(4) | 0.00(4) | - | 0.40(3) | 0.45(4) | 0.45(4) | - |
| | 1000 | 0.65(3) | 0.55(1) | 0.65(3) | - | 0.00(3) | 0.00(1) | 0.00(3) | - | 0.65(3) | 0.55(1) | 0.65(3) | - |
| | 2000 | 0.65(4) | 0.6(3) | 0.65(4) | - | 0.00(4) | 0.00(3) | 0.00(3) | - | 0.65(4) | 0.60(3) | 0.65(4) | - |
| *Case 2* | 500 | 0.35(2) | 0.50(4) | 0.40(3) | - | 0.00(2) | 0.05(4) | 0.10(3) | - | 0.46(2) | 0.58(4) | 0.53(3) | - |
| | 1000 | 0.55(3) | 0.65(3) | 0.60(3) | - | 0.00(3) | 0.00(3) | 0.00(3) | - | 0.70(3) | 0.77(3) | 0.73(3) | - |
| | 2000 | 0.75(7) | 0.80(7) | 0.75(7) | - | 0.00(7) | 0.00(7) | 0.00(7) | - | 0.83(7) | 0.9(7) | 0.83(7) | - |