We are grateful to the reviewers for the insightful comments and suggestions. Please see below for our response.

To R1: Q1. with randomly generated graphs and more variables.: Following your suggestion, we have generated graphs randomly with different numbers of latent variables. See Table 1 below for the results. Will include the results in the paper. Q2. "have an example after definition 1": Thanks for the suggestion. Will do it. Q3. “further explain Theorem 1 and Proposition 1.”: Thanks. We will give more explanations about Theorem 1, as well as Prop. 1., in light of the example in Fig. 1. Specifically, Prop. 1 inspires a unified method to handle causal relations between latent variables and those between latent and observed variables; see the discussion in Section 6. Q4. "The example about exogenous set, S1 and S2 in lines 166-172.": In the first case, S_1 is an exogenous set relative to variable set S_2, where $S_1 = \{L_1\}$ and $S_2 = \{L_3, L_4\}$. In the second case, S_1 is not an exogenous set relative to variable set S_2, where $S_1 = \{L_2, L_3\}$ and $S_2 = \{L_3, L_4\}$. Will give more details. Q5. “In Definition 1 (A1), sufficient to say that no $\mathbf{X}$ should be a parent of any latent variable as opposed to saying no $\mathbf{X}$ should be an ancestor of any latent variable?”: Yes, they are equivalent here. We followed the definition in [12] and will give this interpretation. Q6. “the notation $2Dim(\mathbf{L})$”: $2Dim(\mathbf{L})$ means 2 times the dimension of $\mathbf{L}$. Will clarify it.

To R2: Q1. "everything can be transformed into something that looks Gaussian": Yes we agree; however, when such transforms are applied, the relationships between the transformed variables might not be linear. We think the Gaussian and non-Gaussian methods are complementary and have their own strengths. Because Gaussian methods only use the second order statistics, they have wider applicability. Non-Gaussian methods can provide more information of the structure, in light of the high order statistics. In practice, one may decide which method to apply first based on whether the data are linear and non-Gaussian (e.g., as seen from the scatter plots of the variables). Q2. "multivariate Gaussian distributions are strictly excluded in these works.": Yes. In the multivariate Gaussian case, one can apply traditional Tetrad-based methods, although there is no additional structural information informed by non-Gaussianity. Q3. "O-ICA is easy to get stuck in local optima...": We found that O-ICA is easy to get stuck in local optima, unless the underlying sources are very sparse. This was also reported in publications “Discovering unconfounded causal relationships using linear non-gaussian models" (by Entner, et al., 2011) and “ParceLiNGAM: A causal ordering method robust against latent confounders" (by Tashiro, et al., 2014). However, to avoid possible confusion, we will remove this statement. Q4. "what may not be the case?": Here, we mean that focusing on causal relationships between observed variables alone may not be enough, and the causal structure over latent variables might be very informative. For example, in some cases, the measured variables (such as questionnaire answers) may not loyally reflect the underlying variables of interest and the interesting causal process is over the latents. Q5. "the exact differences between the Triad constraints and the present work should be highlighted.": Thanks for your suggestion. The Triad condition can be seen as a restrictive, special case of the GIN condition, where $Dim(Y) = 2$ and $Dim(Z) = 1$. We will make it explicit.

To R3: Q1. "Algorithm is complete?": Yes, it is complete, as implied by Theorem 3 and Proposition 2 (for step 1) and Theorem 4 and Proposition 3 (for step 2). Will make it explicit. Q2. "is it Dim(S) or Dim(P)?": Thanks for pointing out the typo. It should be Dim(P). Has been corrected.

To R4: Q1. "State assumptions earlier...motivated with real world examples": The assumptions were explicitly given as A1-A4, in the definition of LiNGLaM. Following your suggestion, we will include the assumptions in Abstract and give illustrative real examples. Q2. "A3 does not seem much milder than Tetrad": Compared to Tetrad-based methods, our proposal involves less restrictive structural assumptions but produces stronger results. For instance, under the non-Gaussianity assumption, the graph in the Figure 1 can be recovered by the proposed method, but not by Tetrad-based methods. Q3. "The paper by Anandkumar et al.": Thanks for the suggested this interesting work. This paper makes use of non-Gaussainity of latent variables and was innovative; we will discuss its connection to and difference from our model. We are doing empirical comparisons with it. Q4. "Algorithm 2 returns a causal graph as opposed to an order.": Algorithm 2 indeed returns a causal order of the latent variables. Based on the order, one may directly obtain the causal structure by further estimating the linear coefficients and pruning redundant edges; please see the discussion in lines 333-337. Q5. "Additional experiment where DAGs are sampled randomly.": Thanks for the helpful suggestion. Table 1 below gives the results with randomly generated graphs. Q6. “error of GIN is always the lowest?”: Many thanks for your careful observation. The *Mismeasurements* are higher in Case 3 when the sample size is small (N=500). We will update our claim and explain why this happens. Q7. "Line 244: the boxes indicate?": The boxes indicate the elements of the root variable set $\{L_1, L_2\}$. Will explain it in the paper.

Table 1: Results with different numbers of variables and randomly generated graphs (sample size=2000).

Number of variables (latent variables)	Latent omission	Latent commission	Mismeasurements	Correct-ordering rate
15(5)	0.02(1)	0.00(0)	0.00(0)	0.90
30(10)	0.09(3)	0.05(3)	0.04(3)	0.85
60(20)	0.15(6)	0.12(6)	0.10(6)	0.79