We would like thank the reviewers for the detailed comments and suggestions, which we believe will improve the manuscript.

**Reviewer #1.** **1.** The sample complexity lower bound proved in Thm. C.1 shows that for large values of $p$, $n$ grows like $\log m$ (please see the inequality below line 543). Consequently, if $m$ is a growing function of $p$, no method will exhibit blessing of dimensionality. **2.** We would wish to write down the empirical version of the final objective (3) as a function of the inputs and parameters nicely in a single line. However, it requires introducing notation for samples and adding many repetitive equations for estimating $R_{j,i}, B_{j,i}, r_i$, and $\nu_{X_i|Z}$. **4.** We agree that Gaussian log-likelihood is not the best evaluation metric when the data is non-Gaussian. However, we would like to note that in our experiments most of the baselines, such as glasso, latent variable glasso, [sparse] PCA, are derived under Gaussian assumption. **7.** The "law of total variance", should be "law of total *covariance*". We apologize for the confusion, and will fix this in the camera-ready version. **10.** We will add definitions of quantities such as $H(X), I(X:Y), TC(X), TC(X|Z)$ in the notation paragraph of section 2. **5,6,8,9,11-13.** These flaws and issues will be fixed in the camera-ready version.

**Reviewer #2.** **1.** With the experiments on synthetic data generated from a modular latent factor model, we primarily wanted to demonstrate that the proposed method exhibits blessing of dimensionality. Another goal was to show that, perhaps unsurprisingly, linear CorEx outperforms other methods in covariance estimation. We decided to have this latter result in the main text, since it also demonstrated how the performance gap changes with the number of samples and sparsity level of the ground truth covariance matrix. **2.** We agree that the set of baselines we considered in the stock market experiments probably does not contain the best methods for that specific task. The goal of that experiment was not to show state-of-the-art results on stock market data, but to compare the proposed method with other general covariance estimation methods on a useful real-world dataset. Nevertheless, we look forward to convey this more clearly and cite relevant literature. **3.** In the experiment on fMRI data, the goal is the demonstrate that linear CorEx scales for voxel-level analysis of brain data, while producing meaningful results. We believe the potential of our method for analysing brain data is yet to be explored, with more careful experiments that will contain multiple patients or multiple sessions. **4.** Regarding to the comment on correlation explanation methods proposed by authors of [40] and [41], we will elaborate on the relation of our method with those methods.

**Reviewer #3.** We are happy to fix misleading aspects of the manuscript and improve its clarity. **1.** We will make it explicit that is the stepwise computational complexity of the method that is linear w.r.t the number of variables. **2.** We agree that the exact goal of the approach – finding a latent factor model that is close to being modular – is not stated clearly. We apologize for this and will fix this. **3.** As the reviewer correctly points out, when parameterizing $p_W(z|x)$ we make an implicit assumption that $Z$s are conditionally independent given $X$. We make this assumption since it simplifies the further derivations and is respected by modular latent factor models. This note will be added in the camera-ready version.

Below are our clarifications to your questions. Please note that these clarifications and your suggestions will be added in the camera-ready version.

36, 208 Modular latent factor models have block diagonal covariance matrices. Additionally, each block is a diagonal plus rank-one matrix. To see this, consider a block where $Z$ is the parent and $X_1, X_2, \ldots, X_p$ are the children with respectively $\rho_1, \rho_2, \ldots, \rho_p$ Pearson correlation coefficients with $Z$. Then, $\text{Cov}[X_i, X_{j\neq i}] = \sigma_i \sigma_j \rho_i \rho_j$, where $\sigma_i^2 = \text{Var}[X_i]$. We decided to mention this property of covariance matrices of modular latent factors to demonstrate which types of covariance matrices are preferred by the objective of our method. We though this would be helpful for comparing our method with [latent variable] glasso.

36-38 The intuition is that in those types of data variables can be divided into clusters, where each cluster is governed by a few latent factors and latent factors of different clusters are close to be independent.

40 Learning becomes easier in terms of structural error.

145 We measure the gain by log-likelihood.

246-247 We choose the number of factors according to their log-likelihood score on the validation data.

252-253 Empirical covariance matrices in those cases are not invertible and computing negative log-likelihood becomes impossible.

271 Session 014 is the first publicly available session of that dataset.

272-273 Spatial smoothing intensifies correlations between nearby voxels, helping our model to pick-up the spatial information faster. Without spatial smoothing the training is unstable and we suspect that more samples are needed to train the model.