A Technical Lemma

In Theorem 2, we use four matrix norm inequalities that are less common in the literature. In this section, we prove them in detail.

Lemma 8. *For* $\mathbf{A} \in \mathbb{R}^{N \times N}$*, the following conditions hold:*

$$\begin{array}{ll} \text{i.} & \|\mathbf{A}\|_2 \leq \sqrt{N}\|\mathbf{A}\|_{\infty,2} \\ \text{ii.} & \|\mathbf{A}\|_2 \leq N\|\mathbf{A}\|_{\infty,1} \\ \text{iii.} & \|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_{1,2} \\ \text{iv.} & \mathbf{A} \succ \mathbf{0} \Rightarrow \|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_{1,\infty} \end{array} \tag{14}$$

Proof. Claim i follows from $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_{\mathfrak{F}} \leq \sqrt{N}\|\mathbf{A}\|_{\infty,2}$. The last inequality is equivalent to $\|\mathbf{A}\|_{\mathfrak{F}}^2 \leq N\|\mathbf{A}\|_{\infty,2}^2 \Rightarrow \sum_{n_1 n_2} a_{n_1 n_2}^2 \leq N \max_{n_1} (\sum_{n_2} a_{n_1 n_2}^2)$. Let $|c_{n_1}| = \sum_{n_2} a_{n_1 n_2}^2$, we get $\sum_{n_1} |c_{n_1}| \leq N \max_{n_1} |c_{n_1}|$. This is equivalent to $\|\mathbf{c}\|_1 \leq N\|\mathbf{c}\|_\infty$, and we prove our claim.

Claim ii follows from $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_{\mathfrak{F}} \leq \|\mathbf{A}\|_1 \leq N\|\mathbf{A}\|_{\infty,1}$. The last inequality is equivalent to $\sum_{n_1 n_2} |a_{n_1 n_2}| \leq N \max_{n_1} (\sum_{n_2} |a_{n_1 n_2}|)$. Let $|c_{n_1}| = \sum_{n_2} |a_{n_1 n_2}|$, we get $\sum_{n_1} |c_{n_1}| \leq N \max_{n_1} |c_{n_1}|$. This is equivalent to $\|\mathbf{c}\|_1 \leq N\|\mathbf{c}\|_\infty$, and we prove our claim.

Claim iii follows from $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_{\mathfrak{F}} \leq \|\mathbf{A}\|_{1,2}$. The last inequality is equivalent to $\sqrt{\sum_{n_1 n_2} a_{n_1 n_2}^2} \leq \sum_{n_1} \sqrt{\sum_{n_2} a_{n_1 n_2}^2}$. Let $c_{n_1}^2 = \sum_{n_2} a_{n_1 n_2}^2$, we get $\sqrt{\sum_{n_1} c_{n_1}^2} \leq \sum_{n_1} \sqrt{c_{n_1}^2} = \sum_{n_1} |c_{n_1}|$. This is equivalent to $\|\mathbf{c}\|_2 \leq \|\mathbf{c}\|_1$, and we prove our claim.

Claim iv further assumes that $\mathbf{A}$ is symmetric and positive definite. In this case the spectral radius is less than or equal to any induced norm, specifically the $\ell_{\infty,1}$-norm also called the *max absolute row sum* norm. The inequality we want to prove is $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_{\infty,1} \leq \|\mathbf{A}\|_{1,\infty}$. The last inequality is equivalent to $\max_{n_1} (\sum_{n_2} |a_{n_1 n_2}|) \leq \sum_{n_1} (\max_{n_2} |a_{n_1 n_2}|)$, which follows from the Jensen's inequality. □

B Additional Experimental Results

In what follows, we test the performance of our methods with respect to edge density and the proportion of connected nodes. The following results complement Fig.2 which reported KL divergence between the recovered models and the ground truth for the "low variance confounders" regime. Fig.5 and 6 show the ROC curves and KL divergence between the recovered models and the ground truth for the "high variance confounders" regime. Our $\ell_{1,2}$ and $\ell_{1,\infty}$ methods recover ground truth edges better than competing methods (higher ROC) when edge density among connected nodes is moderate (0.5) to high (0.8), regardless of the proportion of connected nodes. Our proposed methods get similarly good probability distributions (comparable KL divergence) than the other techniques. In the "low variance confounders" regime reported in Fig.7 and Fig.2, our proposed methods produce better probability distributions (lower KL divergence) than the remaining techniques. The behavior of the ROC curves is similar to the "high variance confounders" regime.

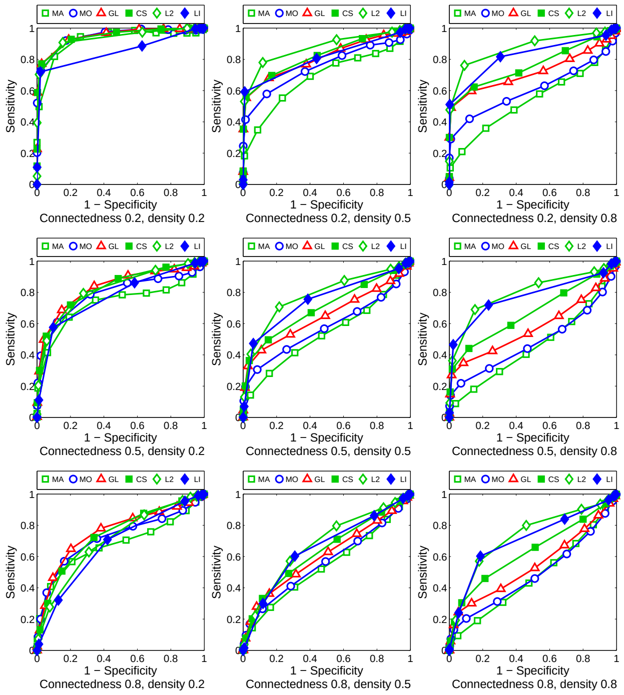

Figure 5: ROC curves for structures learnt for the "high variance confounders" regime ($N = 50$ variables, different connectedness and density levels). Our proposed methods $\ell_{1,2}$ (L2) and $\ell_{1,\infty}$ (LI) recover the ground truth edges better than Meinshausen-Bühlmann with AND-rule (MA), OR-rule (MO), graphical lasso (GL) and covariance selection (CS), when the edge density among the connected nodes is moderate (center) to high (right).

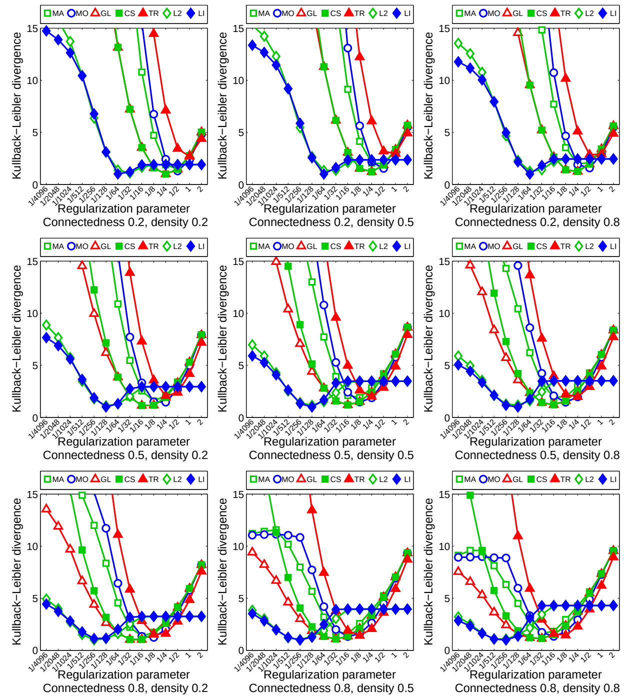

Figure 6: Cross-validated KL divergence for structures learnt for the "high variance confounders" regime ($N = 50$ variables, different connectedness and density levels). Our proposed methods $\ell_{1,2}$ (L2) and $\ell_{1,\infty}$ (LI) produce similarly good probability distributions than Meinshausen-Bühlmann with AND-rule (MA), OR-rule (MO), graphical lasso (GL), covariance selection (CS) and Tikhonov regularization (TR).

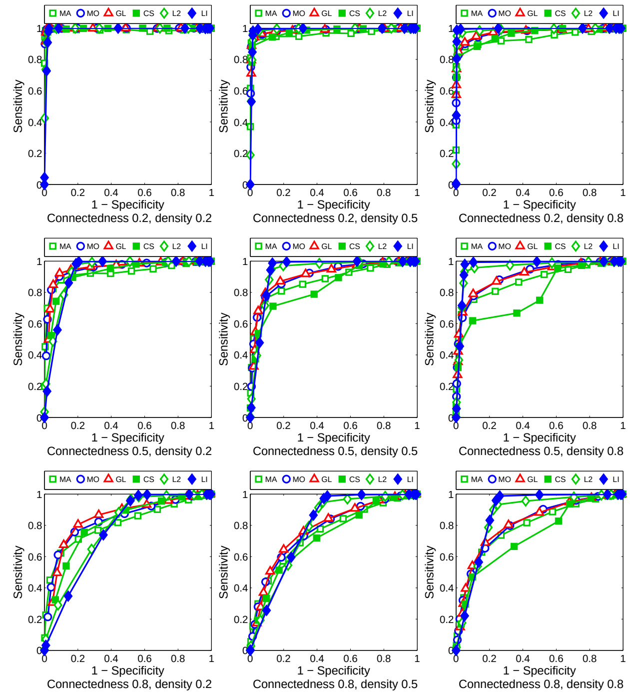

Figure 7: ROC curves for structures learnt for the "low variance confounders" regime ($N = 50$ variables, different connectedness and density levels). Our proposed methods $\ell_{1,2}$ (L2) and $\ell_{1,\infty}$ (LI) recover the ground truth edges better than Meinshausen-Bühlmann with AND-rule (MA), OR-rule (MO), graphical lasso (GL) and covariance selection (CS), when the edge density among the connected nodes is moderate (center) to high (right).