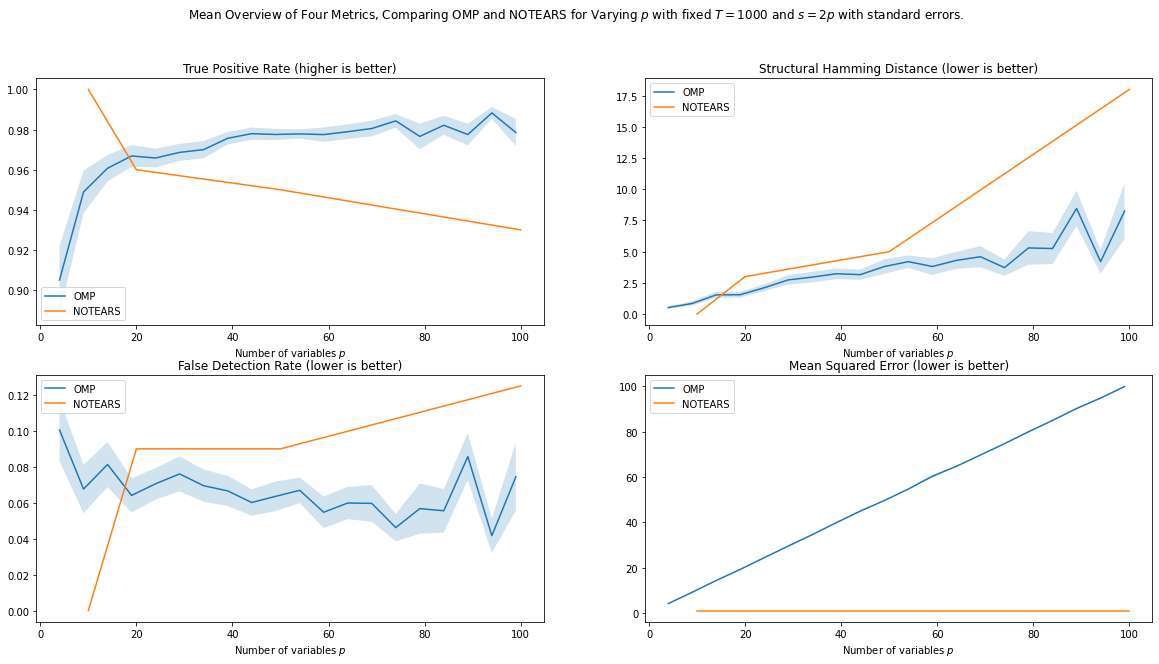
Prep Meeting 23

# Simulation Results

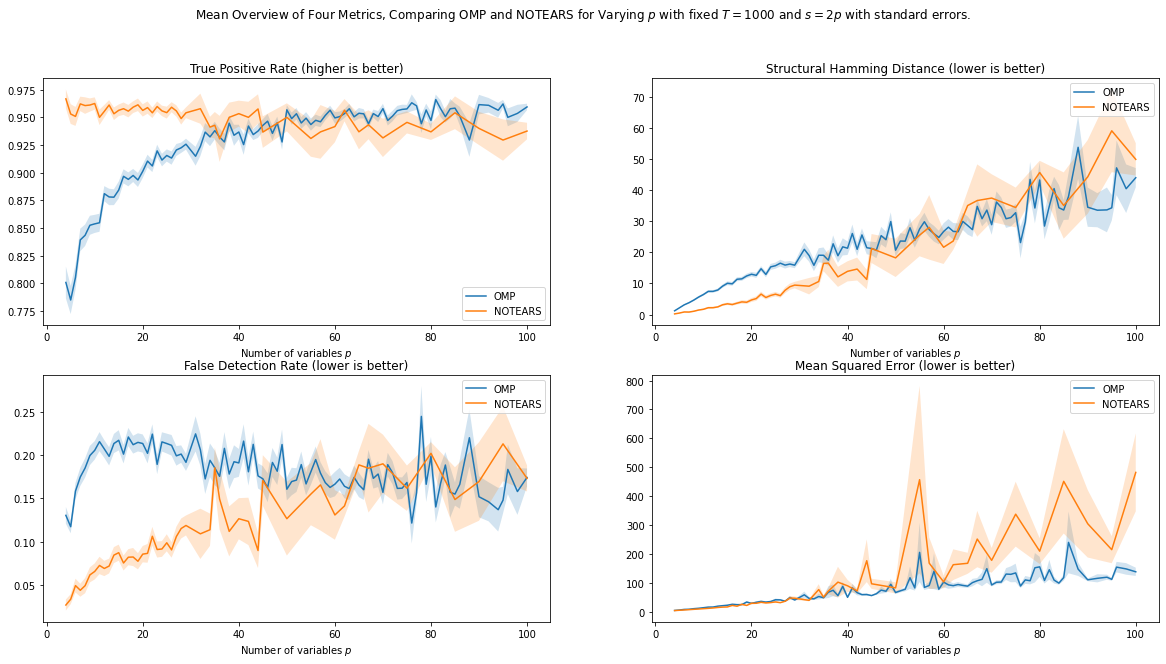
## More Edges

Gaussian Noise, NOTEARS setting, but now with 4p edges rather than 2p edges. This makes the original matrix more dense. Most likely, the greedy method will suffer from this; the sparser the original matrix, the better OMP will fare compare to NOTEARS. Most likely, OMP suffers more from a dense graph than NOTEARS does.

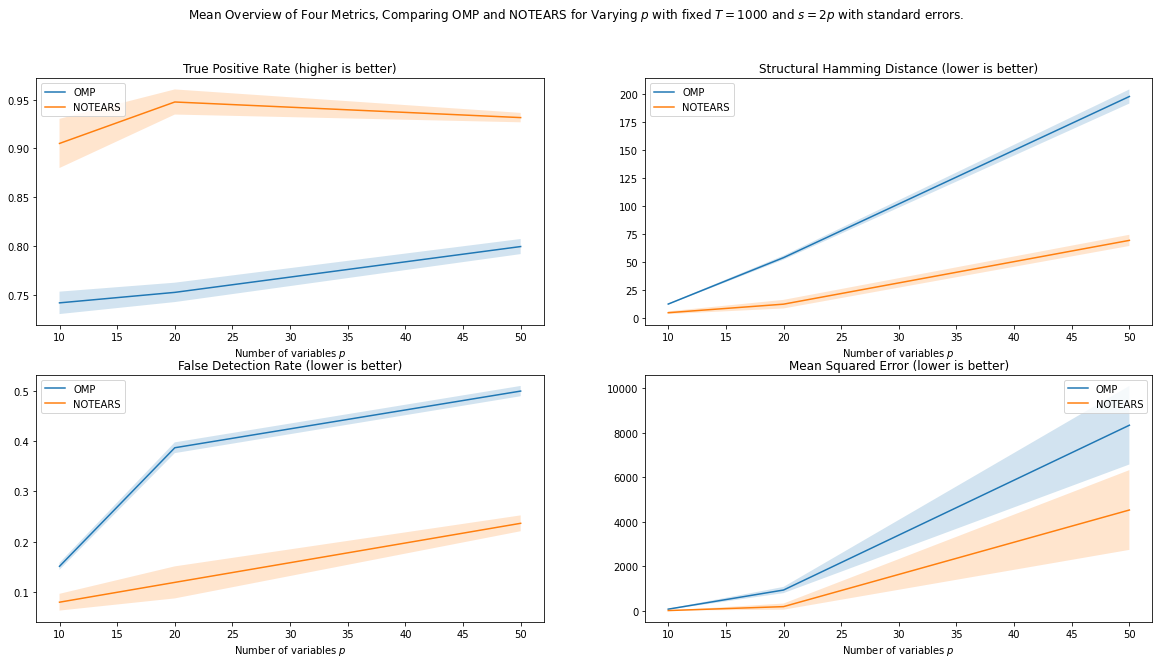
### ER-1, Gaussian noise, 1000 Samples



### ER-2, Gaussian noise, 1000 Samples

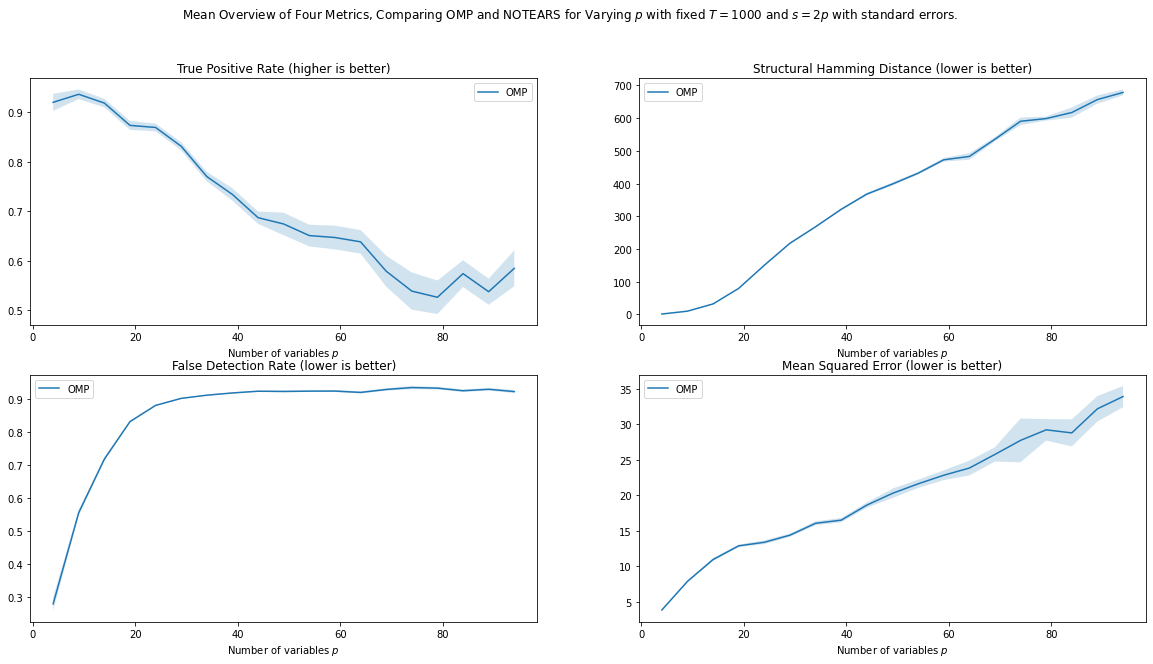


### ER-4, Gaussian noise, 1000 Samples



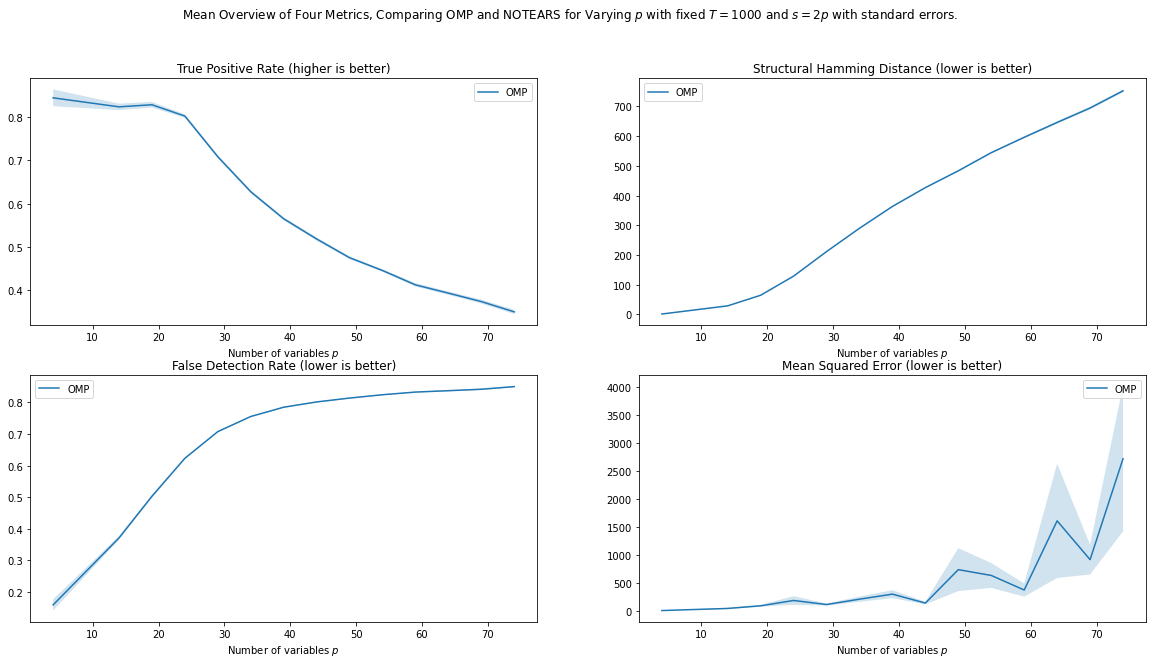
## Different Graph Type

### SF-1, Gaussian Noise, 20 Samples



Remarks: TPR plummets to 0.5, FDR skyrockets to 0.9. SHD also through the roof.

### SF-4, Gaussian Noise, 20 Samples

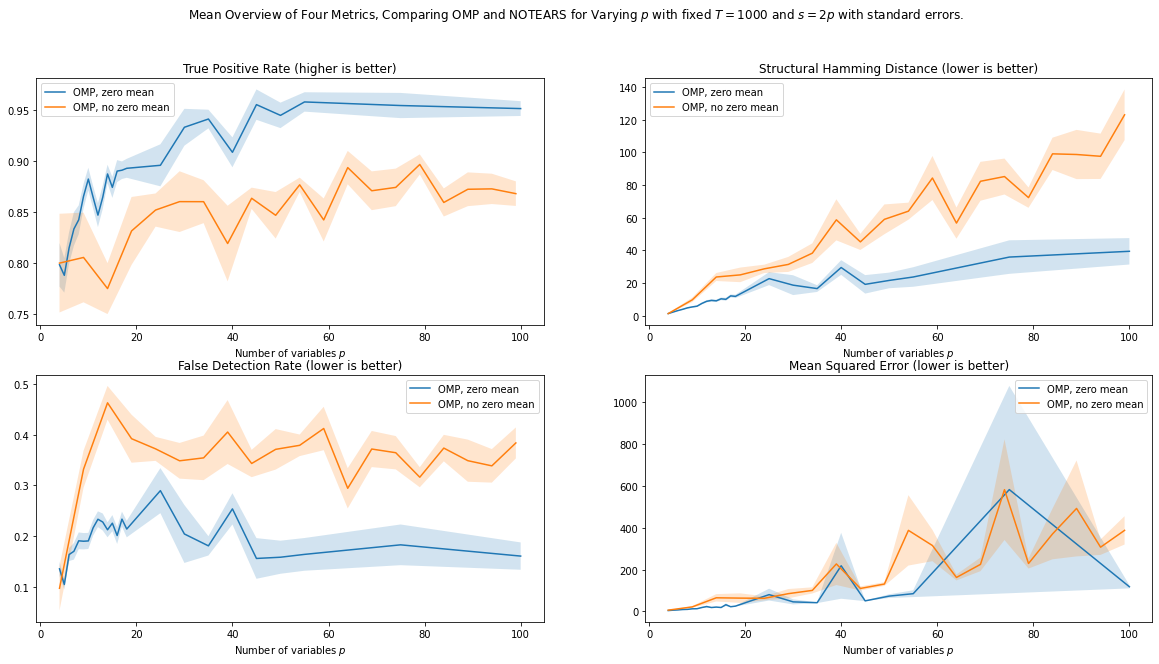


Remark: Same behavior, TPR plummets even more, SHD roughly the same, FDR slightly lower.

## Different Noise Type

Noise Type is also something interesting to investigate. Some methods rely on non-gaussianity (LiNGAM), but it is interesting to see whether OMP is robust against that. Since we just minimize Mean Squared Error, I expect OMP to work well on other noise types. However, we should make sure that we zero – mean the data, just as NOTEARS does.

### ER-2, Exponential Noise, 1000 Samples, With and Without zero mean

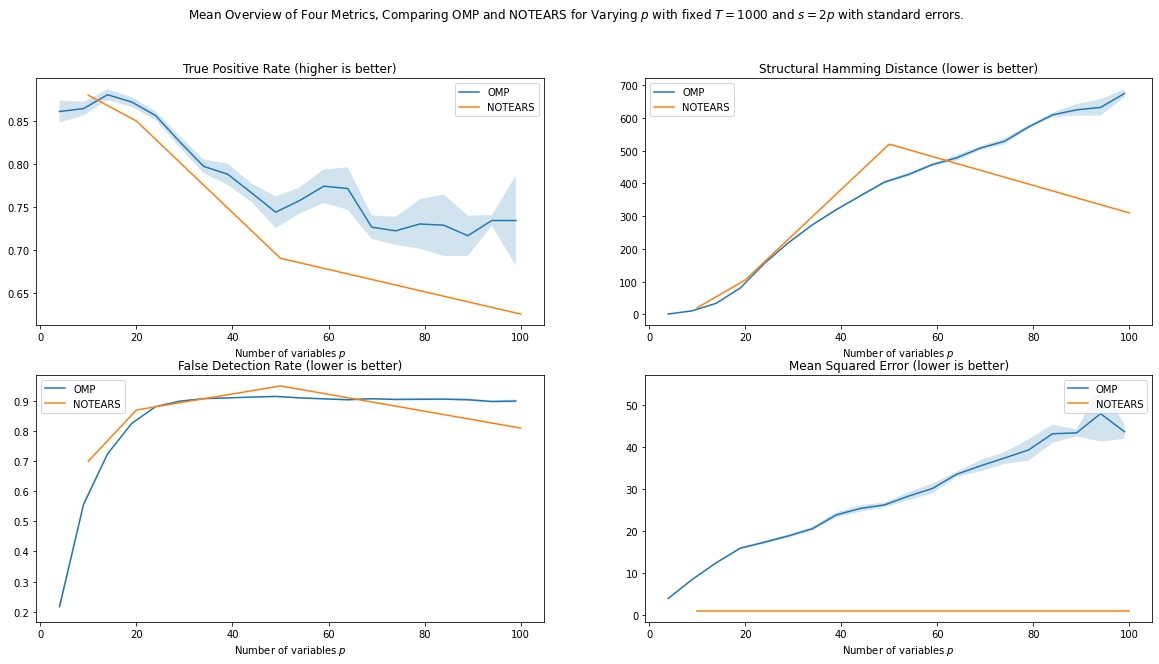


Remark: Zero-meaning the data helps, as the added noise is does not have a zero mean. Results are quite similar, SHD for Exp is much lower, TPR is slightly lower.

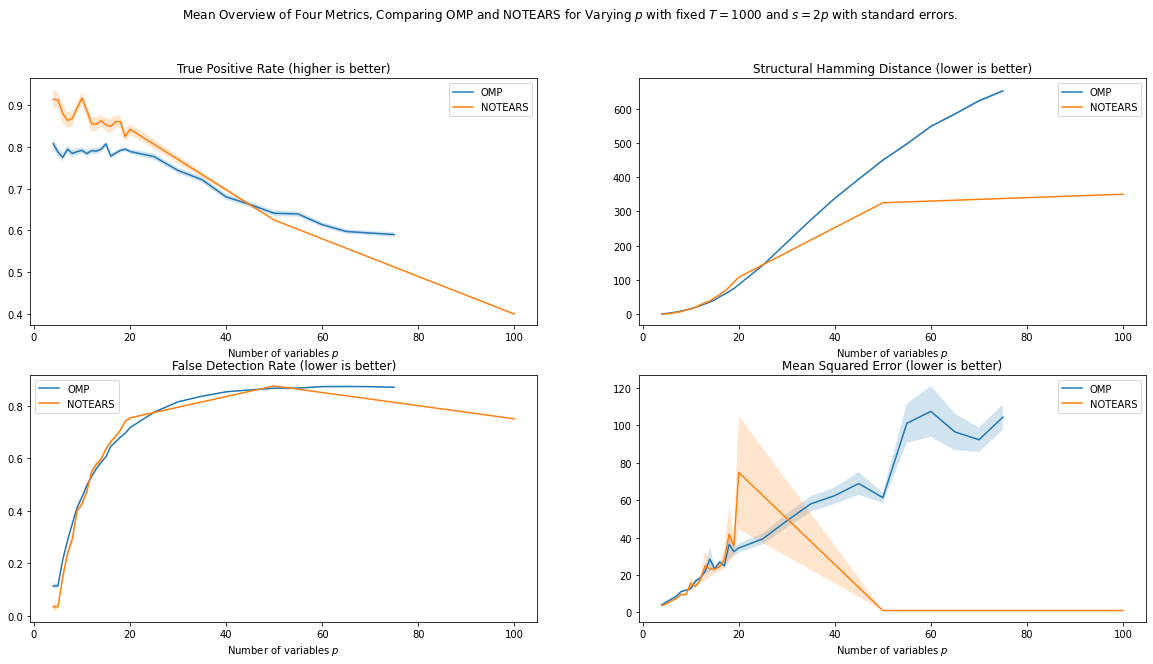
## Different Sample Size

We see that 1000 samples is quite a lot, and then, the problem is much easier. What if we have very few samples, e.g. only 20. Especially when we have 100 nodes, or 10.000 possible edges, fitting those 10.000 parameters using only 20 \* 100 = 2.000 data points will become much more difficult.

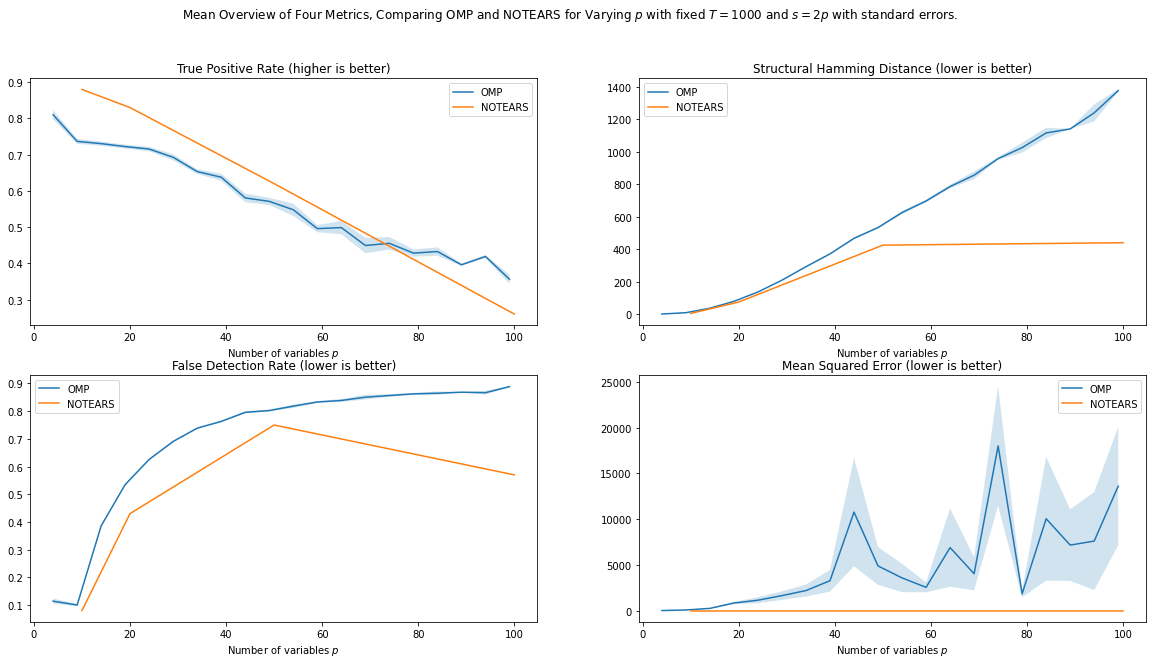
### ER-1, Gaussian noise, 20 Samples



### ER-2, Gaussian noise, 20 Samples



### ER-4, Gaussian noise, 20 Samples



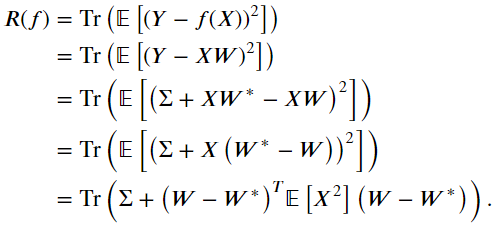
# Regularization VAR

**Small Recap on AUC**

Of course, we can look at some AUC, but for that we need the **true** matrix. A high AUC does indeed show that the regularization method is *capable* of distinguishing between true positives and false positives. *However*, it does not show how *easy* it is for the method to find a “good” TPR-FPR rate. The threshold that corresponds to a “good” TPR-FPR rate may be very small, meaning that the regularization method could be very unstable with respect to its threshold. On the other hand, it could be that there is a wide range of threshold such that the TPR-FPR rate is high, which is also easy to derive. The AUC does not capture this.

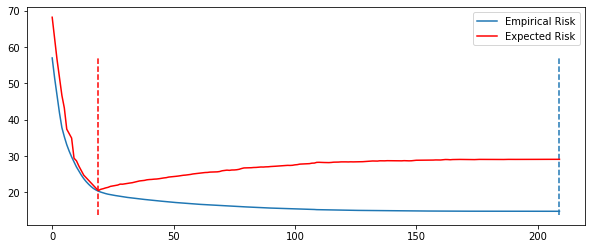
## Cross Validation

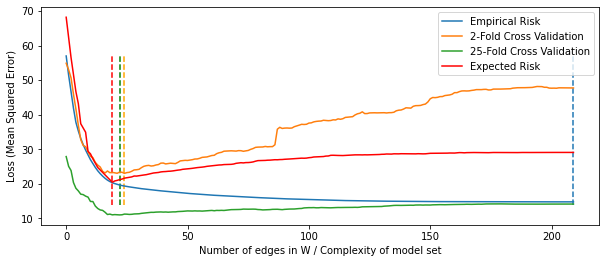
The goal of cross validation is to get a suitable surrogate for the True Risk using only Empirical Risk. Luckily, in our simulations, we know the Expected True Risk. For a VAR model, we know the expected true risk is. The expected excess risk is given by.



Naturally, the **empirical risk** could be used.

## Few Data (20 Samples)





Empirical Risk Minimizer: tpr: **1.0** tnr: 0.50 fpr: 0.90 acc: 0.53 shd: 190.

Naive Risk Minimizer: tpr: **1.0** tnr: 0.97 fpr: 0.35 acc: 0.97 shd: 011.

Two Fold Risk Minimizer: tpr: **1.0** tnr: 0.98 fpr: 0.20 acc: 0.99 shd: 005.

25 Fold Risk Minimizer: tpr: **1.0** tnr: 0.99 fpr: 0.13 acc: 0.99 shd: 003.

Expected Risk Minimizer: tpr: **1.0** tnr: **1.00** fpr: **0.00** acc: **1.00** shd: **000**.

Empirical Minimal Risk: 29.1 => Worst

Naive Minimal Risk: 24.1 => Okay

Two Fold Minimial Risk: 21.6 => Good

25 Fold Minimal Risk: 21.2 => Better

True Minimial Risk: 20.5 => Best

## Much Data (1000 Samples)



Empirical Risk Minimizer: tpr: **1.0** tnr: 0.53 fpr: 0.81 acc: 0.58 shd: 170.

Naive Risk Minimizer: tpr: **1.0** tnr: **1.00** fpr: **0.00** acc: **1.00** shd: **000**.

Two Fold Risk Minimizer: tpr: **1.0** tnr: **1.00** fpr: **0.00** acc: **1.00** shd: **000**.

25 Fold Risk Minimizer: tpr: **1.0** tnr: **1.00** fpr: 0.02 acc: **1.00** shd: 001.

Expected Risk Minimizer: tpr: **1.0** tnr: **1.00** fpr: **0.00** acc: **1.00** shd: **000**.

Empirical Minimal Risk: 20.21 => WORST, BUT NOT BAD

Naive Minimal Risk: 20.07 => OKAY, BUT NOT SUPER

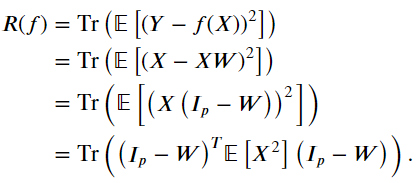
Two Fold Minimial Risk: 20.03 => GOOD

10 Fold Minimal Risk: 20.04 => GOOD

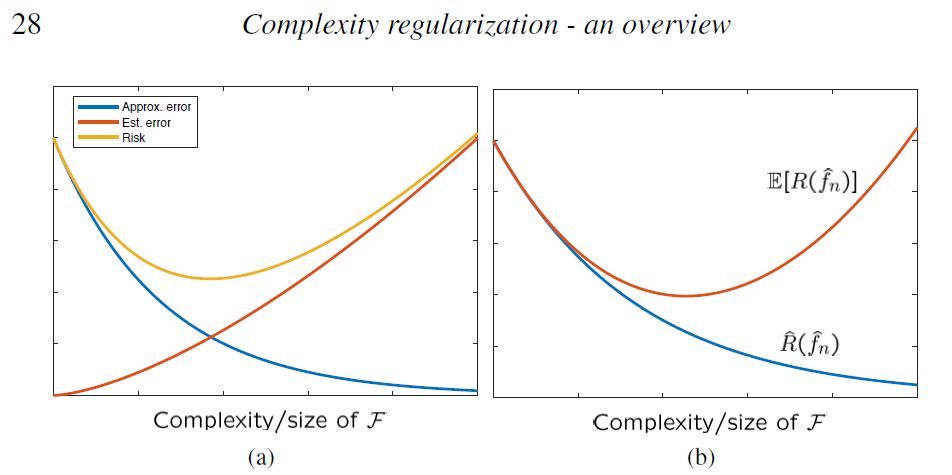
True Minimial Risk: 20.03 => GOOD

# Regularization SEM

Again, here, we know the Bayes risk, and hence we can easily calculate the expected excess risk. The expected excess risk is given by.

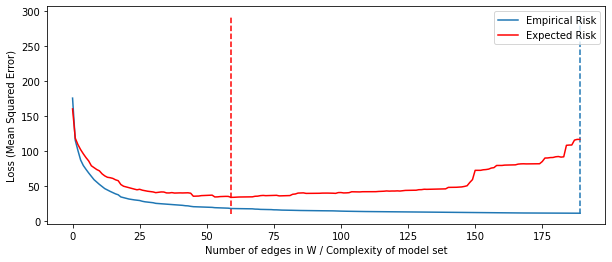


## Cross Validation



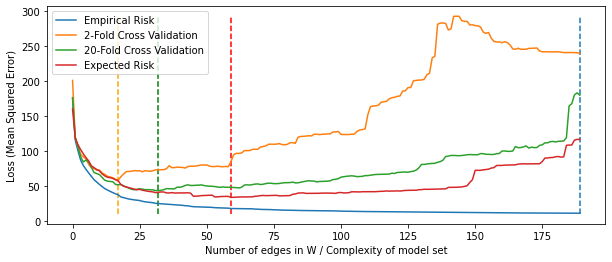
## Few Data (20 Samples)

We indeed see this phenomenon, especially for small sample size:



This is the right plot. For the left plot, we need a way to find the approximation error and the estimation error for varying completities for F, where our complexity lies in the number of non-zero coefficients. For this, we need to find the minimizer W of the class F, which is unfortunately NP-hard to compute, so we cannot really do this decomposition.

However, an approach is to try to get the empirical risk closer to the expected risk, or at least the minima closer together.



We see, a 2-fold cross validation also has an increase in MSE for a more complex model set, but a 20-fold cross validation seems to do better.

Empirical Risk Minimizer: {fdr: 0.82, tpr: **0.88**, fpr: 1.03, shd: 155, nnz: 190}.

Naive Risk Minimizer: {fdr: 0.77, tpr: 0.75, fpr: 0.65, shd: 103, nnz: 128}.

Two Fold Risk Minimizer: {fdr: 0.39, tpr: 0.48, fpr: **0.08**, shd: **29**, nnz:  **41**}.

10 Fold Risk Minimizer: {fdr: **0.06**, tpr: 0.67, fpr: 0.27, shd: 48, nnz: 67}.

Expected Risk Minimizer: {fdr: 0.56, tpr: 0.60, fpr: 0.21, shd: 42, nnz: 55}.

Naive Minimal Risk: 62.9 => BAD, HIGH SHD, THOUGH HIGH TPR.

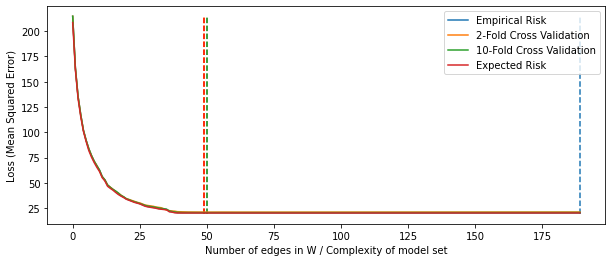
Empirical Minimal Risk: 59.4 => OKAY, HIGH SHD

Two Fold Minimial Risk: 48.0 => GOOD

10 Fold Minimal Risk: 38.8 => BETTER

True Minimial Risk: 35.0 => BEST

## Much Data (1000 Samples)



Empirical Risk Minimizer: {fdr: 0.79, tpr: **0.88**, fpr: 1.00, shd: 155, nnz: 190}.

Naive Risk Minimizer: {fdr: 0.00, tpr: 1.00, fpr: 0.00, shd: 0, nnz: 40}.

Two Fold Risk Minimizer: {fdr: 0.20, tpr: 1.00, fpr: **0.07**, shd: **10**, nnz:  **50**}.

10 Fold Risk Minimizer: {fdr: **0.22**, tpr: 1.00, fpr: 0.07, shd: 11, nnz: 51}.

Expected Risk Minimizer: {fdr: 0.02, tpr: 1.00, fpr: 0.07, shd:  **10**, nnz: 50}.

Empirical Minimal Risk: 20.18 => VERY CLOSE

Naive Minimal Risk: 20.19 => REGULARIZATION INCREASED RISK

Two Fold Minimial Risk: 20.07 => BETTER

10 Fold Minimal Risk: 20.07 => BETTER

True Minimial Risk: 20.07 => BEST