ST443 Machine Learning and Data Mining
Project Report Part.2

36935 (25%)
39290 (25%)
25882 (25%)
30205 (25%)

1 Introduction to two lasso-related approaches

Given a random vector $\mathbf{X} = (X_1, ..., X_p)^T$, each $X_i, i \in \{1, ..., p\}$ can be regarded as a node of the network which consists of those p nodes. Our goal is to estimate the edge set

$$E = \{(j.l) : c_{jl} \neq 0, 1 \leq j, l \leq p, j \neq l\}$$

of such a network, where

$$c_{jl} = \text{Cov}(X_j, X_l | X_k, 1 \leq k \leq p, k \neq j, l)$$

If we assume that $\mathbf{X}$ is multivariate normal, the following two approaches can be applied to estimate edge set E. Notice that to make the sign consistent with the project instruction, we use $\mathbf{X}$ to represent p variables and $\mathbf{x_1}, ..., \mathbf{x_n}$ to represent n samples in the following parts, so $\mathbf{x_i}$ is not a realization of X_i but a realization of $X_1, ..., X_p$.

1.1 Node-wise lasso approach

For each node $j \in V = \{1, ..., p\}$, regress X_j on the remaining variables $X_l, l \in V, l \neq j$, in the form of

$$X_j = \sum_{1 \leq l \leq p, l \neq j} \beta_{jl} X_l + \varepsilon_{jl}$$

and implement lasso approach to estimate $\beta_{jl}, l \in V, l \neq j$. Therefore, with a certain tuning parameter λ, we can estimate E in the following two ways:

$$\hat{E}_{1,\lambda} = \{(j, l) : \text{both } \hat{\beta}_{jl,\lambda} \text{ and } \hat{\beta}_{lj,\lambda} \text{ are nonzero}, 1 \leq j, l \leq p, j \neq l\}$$

$$\hat{E}_{2,\lambda} = \{(j, l) : \text{either } \hat{\beta}_{jl,\lambda} \text{ or } \hat{\beta}_{lj,\lambda} \text{ is nonzero}, 1 \leq j, l \leq p, j \neq l\}$$

1.2 Graphical lasso approach

Denote the covariance matrix of $\mathbf{X}$ as $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ and the inverse covariance matrix as $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$. So by proof, the edge set can be represented by

$$E = \{(j,l) : \Theta_{jl} \neq 0, 1 \leq j,l \leq p, j \neq l\}$$

Then we can minimize the negative log-likelihood with penalty:

$$-\log\det(\mathbf{\Theta}) + \mathrm{tr}(\mathbf{S\Theta}) + \lambda \sum_{j\neq l} \Theta_{jl}$$

to estimate $\mathbf{\Theta}$, where

$$\bar{\mathbf{x}} = \sum_{i=1}^{n} \frac{\mathbf{x_i}}{n},\ \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T$$

And with a certain choice of tuning parameter λ, the estimated edge set is

$$\hat{E}_{3,\lambda} = \{(j,l) : \hat{\Theta}_{jl,\lambda} \neq 0, 1 \leq j,l \leq p, j \neq l\}$$

2 Data generation

We first generate $\mathbf{\Theta} = \mathbf{B} + \delta\mathbf{I_p} \in \mathbb{R}^{p\times p}$ to represents the true edge set $\mathbf{E}$ with its sparsity pattern. Each off-diagonal entry in $\mathbf{B}$ is generated by

$$\mathbf{B}_{ij} \sim 0.5 * Bernoulli(p = 0.1)$$

Then to make $\mathbf{\Theta}$ have unit diagonals, we divide each entry of the matrix with δ. Also, to make sure that $\mathbf{\Theta}$ is a (semi-)positive definite matrix so that $\mathbf{\Sigma} = \mathbf{\Theta^{-1}}$ can perform as a covariance matrix, we iterate δ until all of the eigenvalues of $\mathbf{\Theta}$ are non-negative.

After constructing $\mathbf{\Theta}$, we generate samples $\mathbf{x_1}, \ldots, \mathbf{x_n}$ from a multivariate normal distribution with zero mean and the covariance matrix $\mathbf{\Sigma} = \mathbf{\Theta^{-1}}$. Notice that the sample size n, number of nodes p and the sparsity pattern of $\mathbf{\Theta}$, represented by the probability in *Bernoulli* distribution, will all affect the estimation. Therefore, we fix $n = 500$, $p = 100$ and $Bernoulli(p = 0.1)$ for the next few parts, try other values of those parameters in the last part and present more colorful results in the appendix.

3 ROC curve and compute AUROC

For both node-wise lasso and graphical lasso, we iterate the penalty parameter λ from 0 to 1 with step 0.01 so that for each λ, we have an estimation of $\hat{E}_i,\ i \in \{1,2,3\}$ with the corresponding true positive rate (TPR_λ) and false positive rate (FPR_λ). After plotting each pair of (TPR_λ, FPR_λ) as a dot on a graph with axes of TPR and FPR, we connect those points with a concave step curve to obtain a continuous ROC curve.

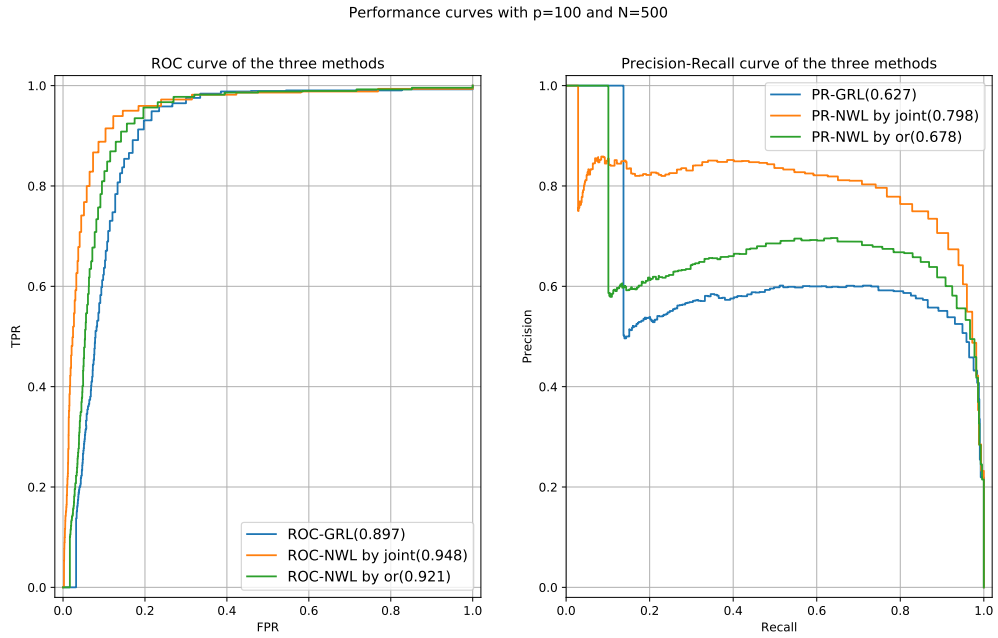


Figure 1: Performance curve with p=100 and n=500

To compute AUROC, we equally divide the horizontal axes of FPR into 1000 intervals. For any interval i, we denote the average value of the part of ROC curve which falls within interval i as y_i. Therefore,

$$\text{AUROC} \approx \frac{1}{1000} \sum_{\text{i=1}}^{1000} \text{y}_\text{i}$$

With $n = 500$ and $p = 100$, AUROC of node-wise lasso 1, node-wise lasso 2 and graphical lasso are 0.948, 0.921 and 0.897.

4 Selection of optimal tuning parameters

Generally, we use cross validation and BIC to select optimal tuning parameters.

4.1 Cross validation

Under the framework of node-wise lasso, for each $X_j, j \in V$, we set a sequence of λ and calculate the mean of test MSE for each λ with cross validation. The λ that minimizes the mean of test MSE will be selected as the tuning parameter of lasso to determine whether $\beta_{jl}, l \in V, l \neq j$ is non-zero.

As for graphical lasso, we directly invoke the GraphicalLassoCV function in Python to calculate $d(\hat{\mathbf{\Sigma}}, \hat{\mathbf{\Theta}})$, which measures how well the estimated inverse covariance fits the data, to conduct cross validation.

4.2 BIC

In terms of node-wise lasso, for each $X_j, j \in V$, we set a sequence of λ to calculate the corresponding BIC of estimation result with each λ.

$$\mathrm{BIC}_\lambda = -2\mathrm{log}L(X_j, \sum_{1\leq l\leq p, l\neq j} \hat{\beta}_{jl}X_l) + df log(n)$$

where

$$\mathrm{log}L(X_j, \sum_{1\leq l\leq p, l\neq j} \hat{\beta}_{jl}X_l) = -\frac{1}{2}\left[ln(2\pi) + ln\left\{\frac{1}{n}\sum_{j=1}^{n}(X_j - \sum_{1\leq l\leq p, l\neq j}\hat{\beta}_{jl}X_l)^2\right\} + 1\right]$$

and *df* is the number of nonzero coefficients. λ that minimizes BIC will be selected.

Similarly, in graphical lasso case, we calculate BIC for each λ with the following formula (Gao et al., 2012):

$$\mathrm{BIC}_\lambda = -n\mathrm{log}\det(\hat{\Theta}) + n\mathrm{tr}(\hat{\Theta}\bar{A}) + \log(n)\sum_{1\leq i<j\leq p} I(\hat{\Theta}_{ij}\neq 0)$$

where

$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})'$$

4.3 Support recovery

The results of the support recovery is presented in the following Table 1. Besides, we also take advantage of

$$\mathrm{F}_1\,\mathrm{score} := \frac{\mathrm{TP}}{\mathrm{TP} + (\mathrm{FP} + \mathrm{FN})/2}$$

and Area:=TPR·(1-FPR) to show the sample performance of each method. The larger they are, the better the performance is.

Generally, in terms of F_1 and Area, BIC performs better than CV as the two indicators of BIC are always larger than that of CV for each model.

For graphical lasso, the model selected by BIC tends to predict the result as positive (nodes j and l are connected) so TP and FP are larger than that of CV. On the contrary, TN and FN of BIC are relatively small. In a word, BIC has larger TPR and FPR.

For node-wise lasso, CV performs a little better when the true result is positive, reflected by TP and FN. However, the FPR of BIC are significantly lower than that of CV, i.e. at this time CV has larger TPR and FPR.

Model	Graphical Lasso		Node-wise Lasso1(joint)		Node-wise Lasso(or)	
Method	CV	BIC	CV	BIC	CV	BIC
TP	325	824	861	776	873	766
FP	200	686	709	259	1315	374
TN	3868	3382	3359	3809	2753	3694
FN	557	58	21	106	9	116
F_1	0.46	0.69	0.7	0.81	0.57	0.76
Area	0.35	0.78	0.81	0.82	0.67	0.79

Table 1: Support recovery by selected optimal tuning parameter

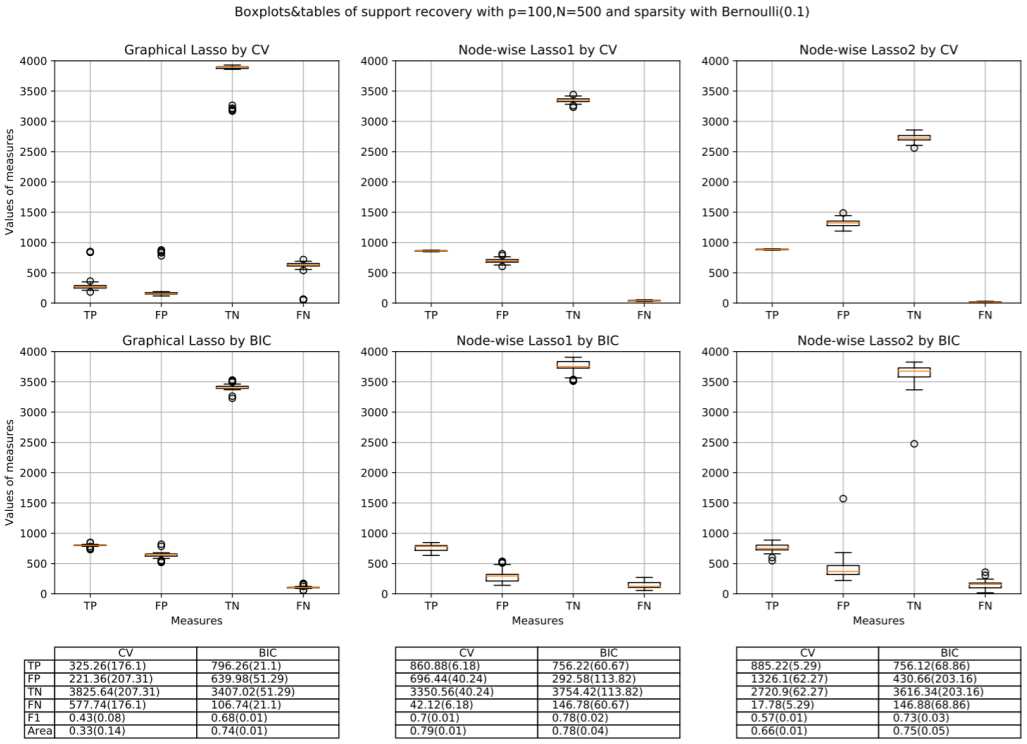


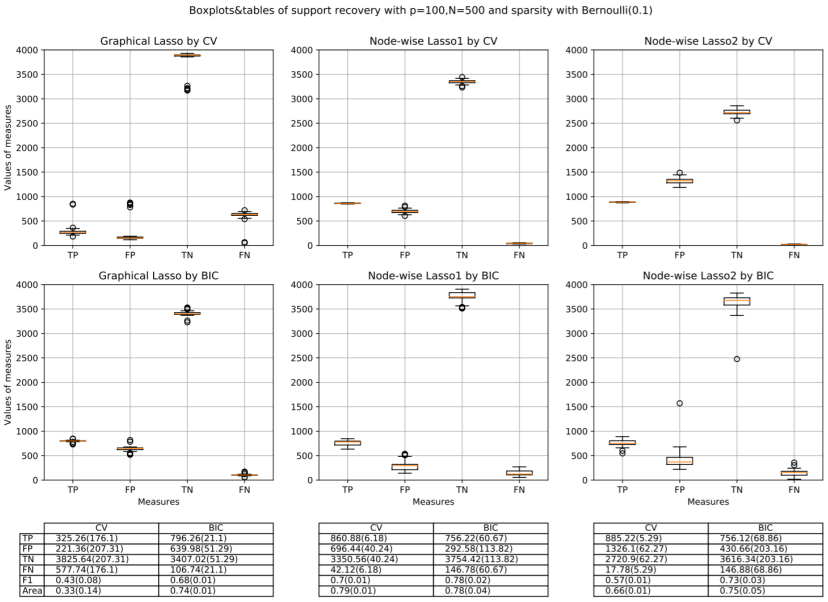
Figure 2: Boxplots and tables of support recovery derived by replication for 50 times

5 Replicate 50 times

The results of the mean and standard error of relevant measure terms derived by replicating for 50 times are presented in the above boxplots and tables. The comparison of the mean of support recovery is quite similar to section 4.3, so we focus on standard error. For graphical lasso, the standard error of CV are larger than that of BIC; while for node-wise lasso, the opposite is true. Within node-wise lasso, the standard error of node-wise lasso1 is smaller than that of node-wise lasso2.

6 Results with different n, p and sparsity

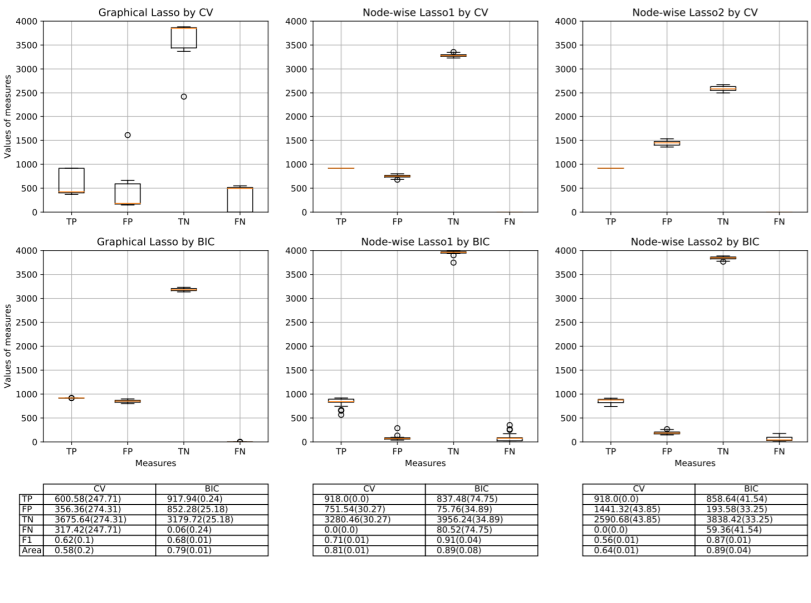
First, we fix p and sparsity structure and compare the results when n is 500 or 2000. When n increases from 500 to 2000, the overall performance is improved in terms of F_1 and Area for each method. For BIC, both TP and TN increase; while for CV, TP increases but TN decreases so that TPR and FPR increase together. Also, the standard error of the support recovery of graphical lasso with CV increase while all of the others decrease.



	CV	BIC
TP	325.26(176.1)	796.26(21.1)
FP	221.36(207.31)	639.98(51.29)
TN	3825.64(207.31)	3407.02(51.29)
FN	577.74(176.1)	106.74(21.1)
F1	0.43(0.08)	0.68(0.01)
Area	0.33(0.14)	0.74(0.01)

	CV	BIC
TP	860.88(6.18)	756.22(60.67)
FP	696.44(40.24)	292.58(113.82)
TN	3350.56(40.24)	3754.42(113.82)
FN	42.12(6.18)	146.78(60.67)
F1	0.7(0.01)	0.78(0.02)
Area	0.79(0.01)	0.78(0.04)

	CV	BIC
TP	885.22(5.29)	756.12(68.86)
FP	1326.1(62.27)	430.66(203.16)
TN	2720.9(62.27)	3616.34(203.16)
FN	17.78(5.29)	146.88(68.86)
F1	0.57(0.01)	0.73(0.03)
Area	0.66(0.01)	0.75(0.05)

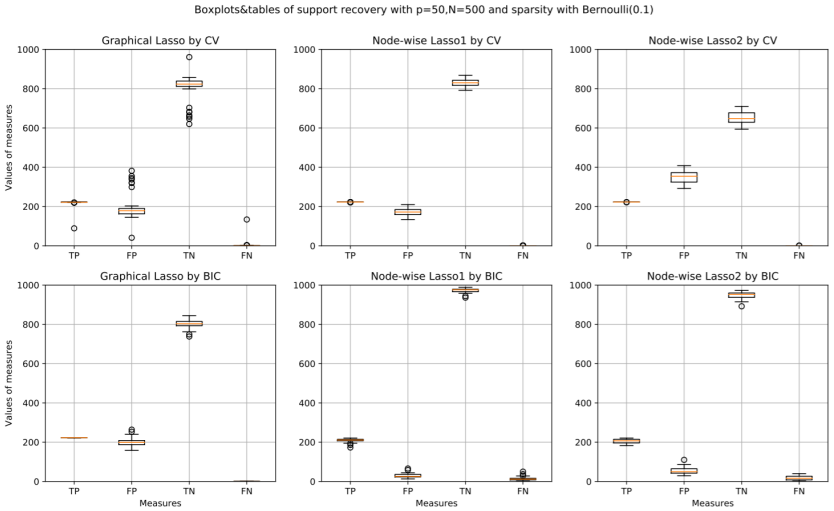


	CV	BIC
TP	600.58(247.71)	917.94(0.24)
FP	356.36(274.31)	852.28(25.18)
TN	3675.64(274.31)	3179.72(25.18)
FN	317.42(247.71)	0.06(0.24)
F1	0.62(0.1)	0.68(0.01)
Area	0.58(0.2)	0.79(0.01)

	CV	BIC
TP	918.0(0.0)	837.48(74.75)
FP	751.54(30.27)	75.76(34.89)
TN	3280.46(30.27)	3956.24(34.89)
FN	0.0(0.0)	80.52(74.75)
F1	0.71(0.01)	0.91(0.04)
Area	0.81(0.01)	0.89(0.08)

	CV	BIC
TP	918.0(0.0)	858.64(41.54)
FP	1441.32(43.85)	193.58(33.25)
TN	2590.68(43.85)	3838.42(33.25)
FN	0.0(0.0)	59.36(41.54)
F1	0.56(0.01)	0.87(0.01)
Area	0.64(0.01)	0.89(0.04)

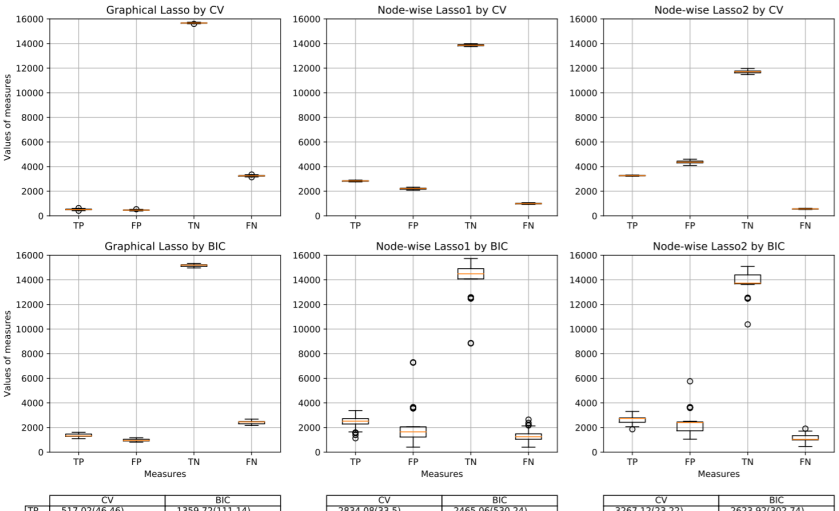
Second, we fix n and sparsity structure and compare the results when p is 50 or 200. Obviously, when p is increased from 50 to 200, the overall performance of each model with either selection methods get worse. It's a quite natural consequence because as p increases, the number of parameters that need to be estimated increase while the information used to estimate them keeps the same.



	CV	BIC
TP	219.38(18.65)	222.52(0.7)
FP	194.72(62.48)	200.52(18.95)
TN	807.28(62.48)	801.48(18.95)
FN	3.62(18.65)	0.48(0.7)
F1	0.69(0.06)	0.69(0.02)
Area	0.79(0.08)	0.8(0.02)

	CV	BIC
TP	222.78(0.46)	209.62(9.1)
FP	171.24(17.84)	29.26(10.33)
TN	830.76(17.84)	972.74(10.33)
FN	0.22(0.46)	13.38(9.1)
F1	0.72(0.02)	0.91(0.02)
Area	0.83(0.02)	0.91(0.03)

	CV	BIC
TP	222.94(0.24)	205.52(10.72)
FP	349.74(27.25)	54.36(16.32)
TN	652.26(27.25)	947.64(16.32)
FN	0.06(0.24)	17.48(10.72)
F1	0.56(0.02)	0.85(0.02)
Area	0.65(0.03)	0.87(0.04)

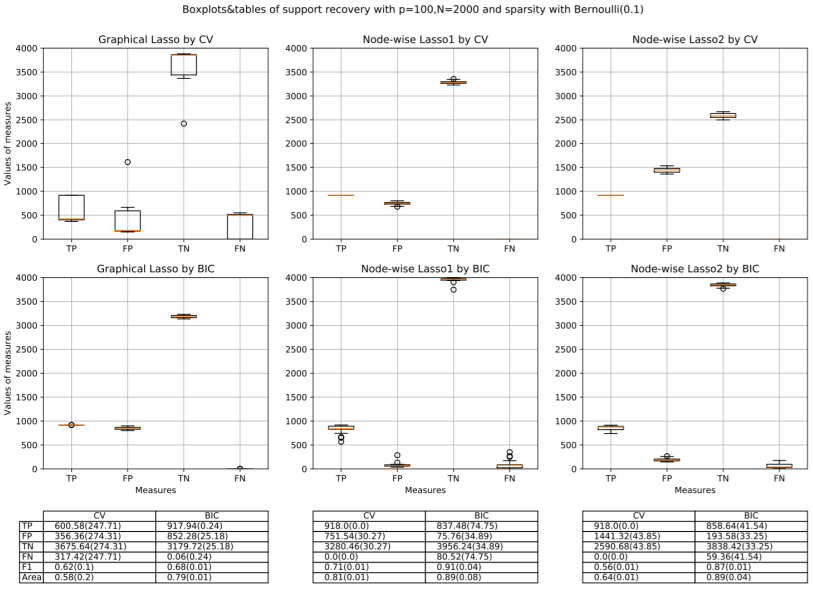


	CV	BIC
TP	517.02(46.46)	1359.72(111.14)
FP	467.06(29.81)	962.84(81.46)
TN	15664.94(29.81)	15169.16(81.46)
FN	3250.98(46.46)	2408.28(111.14)
F1	0.22(0.02)	0.45(0.02)
Area	0.13(0.01)	0.34(0.03)

	CV	BIC
TP	2834.08(33.5)	2465.06(530.24)
FP	2208.18(66.03)	2110.46(1633.43)
TN	13862.82(66.03)	14021.54(1633.43)
FN	994.92(33.5)	1302.94(530.24)
F1	0.64(0.01)	0.59(0.05)
Area	0.64(0.01)	0.56(0.08)

	CV	BIC
TP	3267.12(23.22)	2623.92(302.74)
FP	4362.84(109.66)	2287.24(824.86)
TN	11708.16(109.66)	13844.76(824.86)
FN	561.88(23.22)	1144.08(302.74)
F1	0.57(0.0)	0.61(0.02)
Area	0.62(0.01)	0.59(0.04)

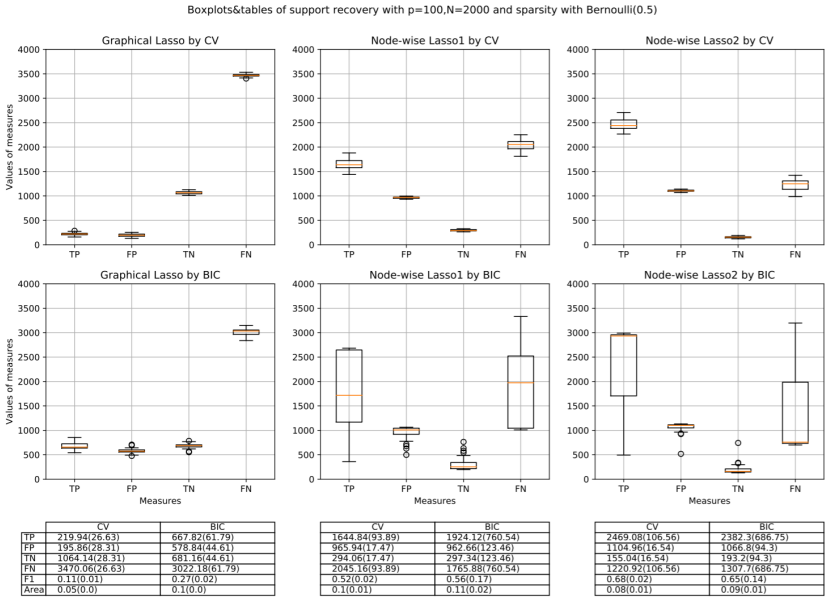
Finally, we keep n and p constant and change the sparsity structure from $Bernoulli(p = 0.1)$ to $Bernoulli(p = 0.5)$. Overall, the performance gets worse as the sparsity level increases. Mean of TN decreases a lot a mean of FN increases. For graphical lasso, the standard error of CV decreases; for node-wise lasso, the standard error of BIC increases and the standard error of TP and FN with CV also increase.



	CV	BIC
TP	600.58(247.71)	917.94(0.24)
FP	356.36(274.31)	852.28(25.18)
TN	3675.64(274.31)	3179.72(25.18)
FN	317.42(247.71)	0.06(0.24)
F1	0.62(0.1)	0.68(0.01)
Area	0.58(0.2)	0.79(0.01)

	CV	BIC
TP	918.0(0.0)	837.48(74.75)
FP	751.54(30.27)	75.76(34.89)
TN	3280.46(30.27)	3956.24(34.89)
FN	0.0(0.0)	80.52(74.75)
F1	0.71(0.01)	0.91(0.04)
Area	0.81(0.01)	0.89(0.08)

	CV	BIC
TP	918.0(0.0)	858.64(41.54)
FP	1441.32(43.85)	193.58(33.25)
TN	2590.68(43.85)	3838.42(33.25)
FN	0.0(0.0)	59.36(41.54)
F1	0.56(0.01)	0.87(0.01)
Area	0.64(0.01)	0.89(0.04)



	CV	BIC
TP	219.94(26.63)	667.82(61.79)
FP	195.86(28.31)	578.84(44.61)
TN	1064.14(28.31)	681.16(44.61)
FN	3470.06(26.63)	3022.18(61.79)
F1	0.11(0.01)	0.27(0.02)
Area	0.05(0.0)	0.1(0.0)

	CV	BIC
TP	1644.84(93.89)	1924.12(760.54)
FP	965.94(17.47)	962.66(123.46)
TN	294.06(17.47)	297.34(123.46)
FN	2045.16(93.89)	1765.88(760.54)
F1	0.52(0.02)	0.56(0.17)
Area	0.1(0.01)	0.11(0.02)

	CV	BIC
TP	2469.08(106.56)	2382.3(686.75)
FP	1104.96(16.54)	1066.8(94.3)
TN	155.04(16.54)	193.2(94.3)
FN	1220.92(106.56)	1307.7(686.75)
F1	0.68(0.02)	0.65(0.14)
Area	0.08(0.01)	0.09(0.01)

References

[1] Gao, X., Pu, D. Q., Wu, Y. and Xu, H.. (2012). Tuning Parameter selection for penalized likelihood estimation of gaussian graphical model, *Statistica Sinica*, 22, p. 1123-1146

[2] *Stata lasso reference annual* (release 17). (2021). College Station, TX: StataCorp LLC

[3] R. G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: with Applications in R.*. (2013). Springer

Appendix 1. Mathematical Derivations

Derive the log-likelihood function of linear regression in section 4.2 BIC
Consider a linear regression model

$$y_i = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon_i = \beta_0 + \mathbf{x_i'}\beta + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

Then the distribution of y_i conditional on $\mathbf{x_i}$ is

$$f(y_i|\mathbf{x_i}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(y_i - \beta_0 - \mathbf{x_i'}\beta)^2}{2\sigma^2})$$

The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\mathbf{x_i})$$

$$\Rightarrow \log L(\theta) = \frac{1}{n}\sum_{i=1}^{n} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y_i - \beta_0 - \mathbf{x_i'}\beta)^2}{2\sigma^2}$$

$$\Rightarrow -2\log L(\theta) = \log(2\pi\sigma^2) + \frac{1}{n\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \mathbf{x_i'}\beta)^2$$

Take derivative of $L(\theta)$ w.r.t. σ^2, and by first order condition $\frac{\partial L(\theta)}{\partial \sigma^2} = 0$:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \mathbf{x_i'}\hat{\beta})^2$$

Therefore,

$$-2\log L(\hat{\theta}) = \log(2\pi) + \log\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \mathbf{x_i'}\hat{\beta})^2\right\} + 1$$

Divide both sides by -2 and then we get the same equation as in section 4.2.

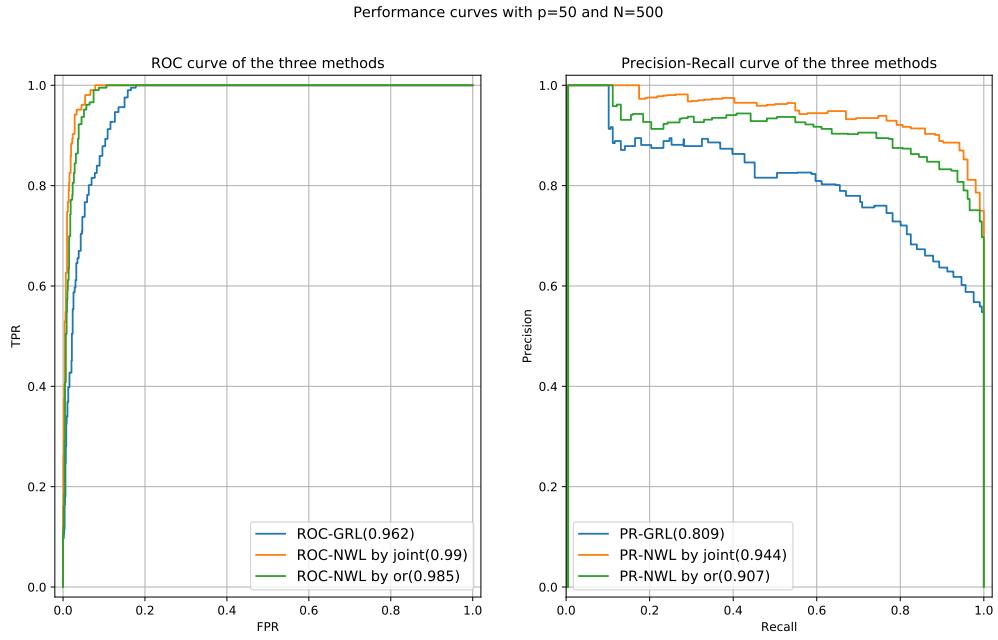
Appendix 2. GitHub link for the python code

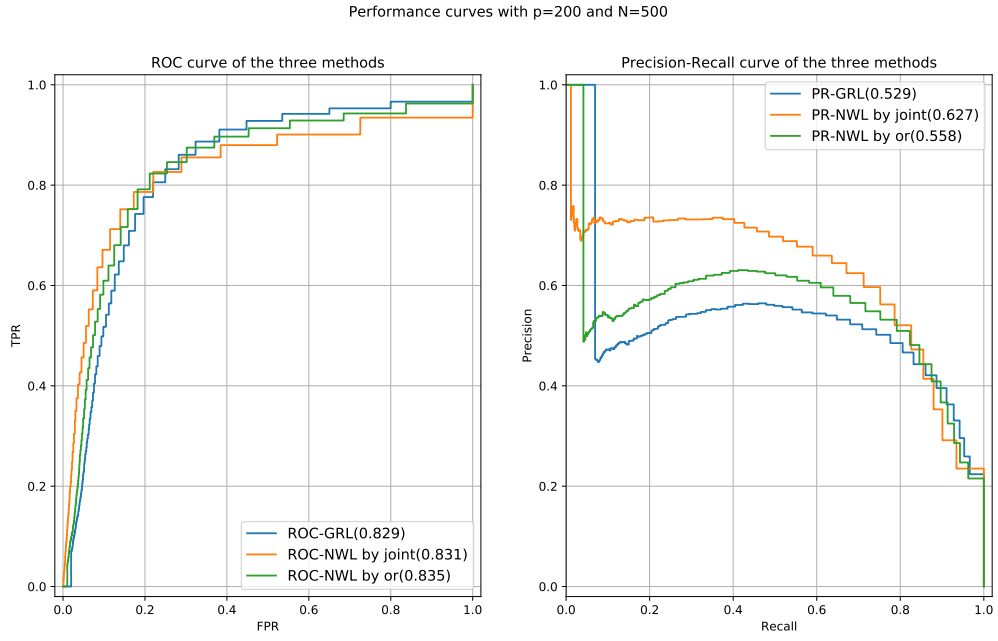
We actually use both Python and R to produce the all the results, and almost all the results produced by the two languages coincide. However, we decide to use Python to produce better plots. Here I include the link to our ST443 project repository, all the codes(Python script and notebook) and figures for this part is in the **Lasso for graphical models** directory.

`https://github.com/ST443project/ST443_project`

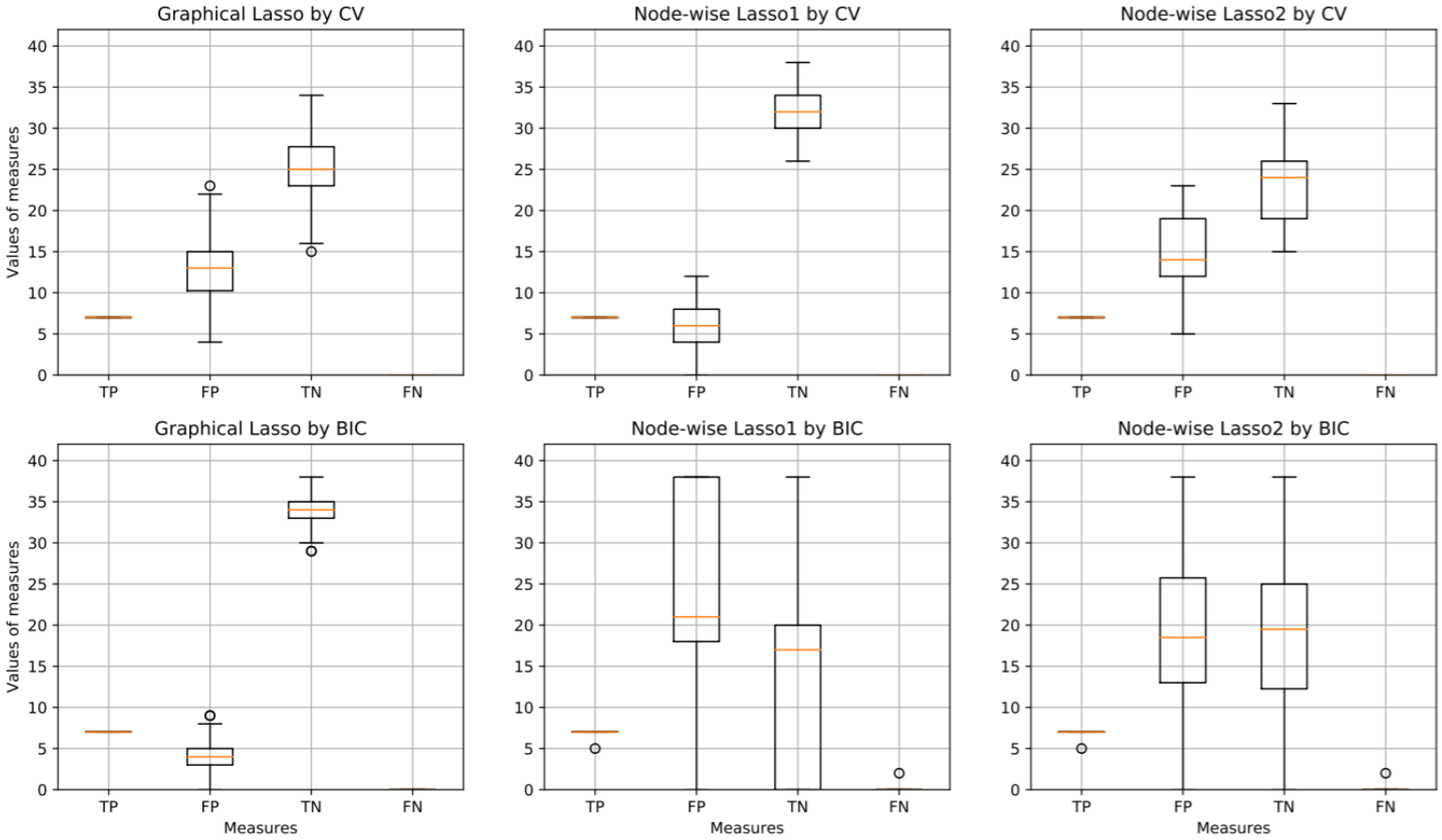
Appendix 3. More results with different parameters

Performance Curve with $p = 50$ and $p = 200$.
Boxplots and tables of support recovery with $n \in \{500, 2000\}$, $p \in \{10, 30, 50, 100, 200\}$, sparsity structure$\in \{0.1, 0.3, 0.5\}$.





Boxplots&tables of support recovery with p=10,N=500 and sparsity with Bernoulli(0.1)

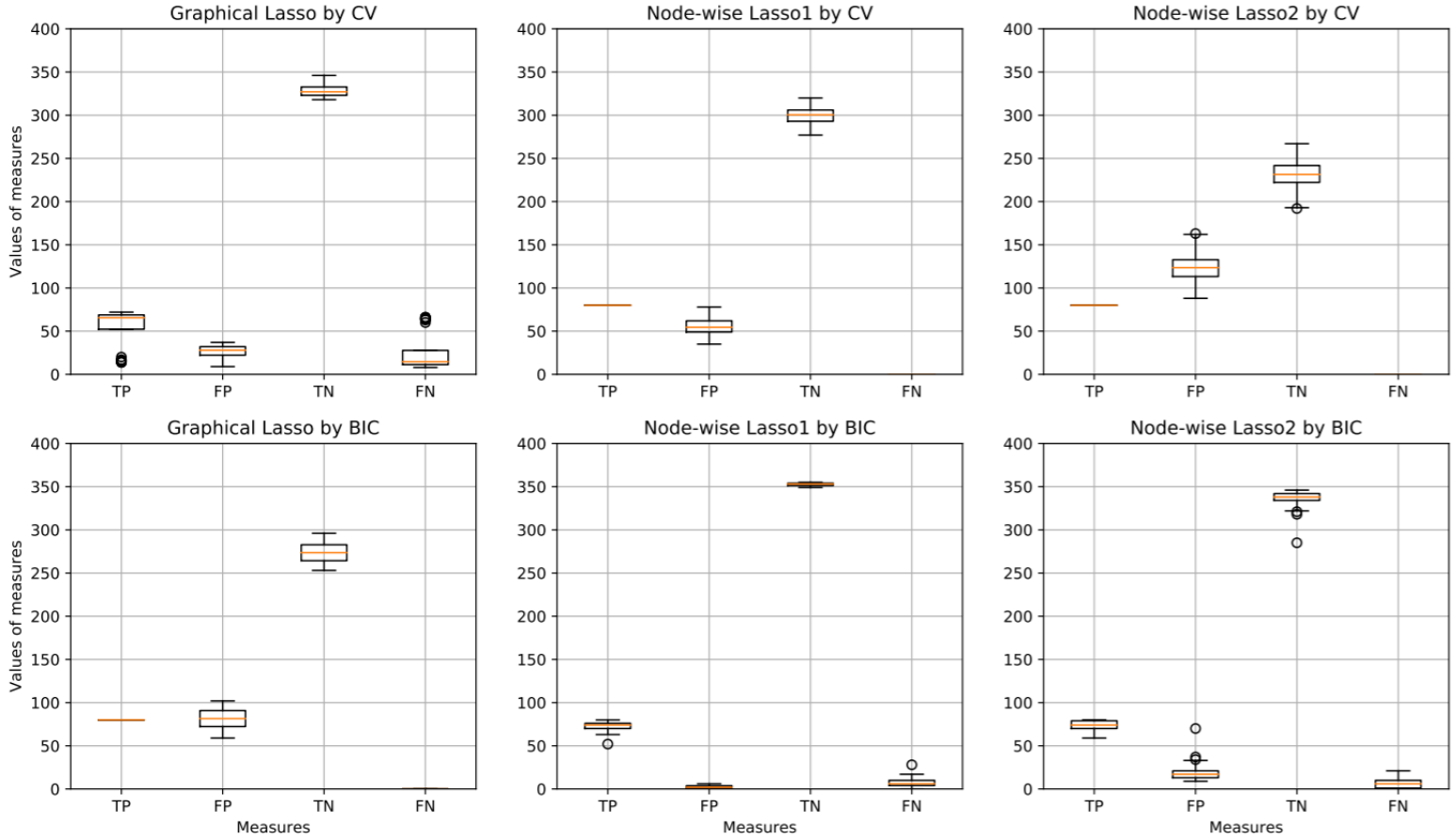


	CV	BIC
TP	7.0(0.0)	7.0(0.0)
FP	13.16(3.98)	4.2(2.14)
TN	24.84(3.98)	33.8(2.14)
FN	0.0(0.0)	0.0(0.0)
F1	0.53(0.08)	0.78(0.09)
Area	0.65(0.1)	0.89(0.06)

CV	BIC
7.0(0.0)	6.96(0.28)
5.78(2.59)	22.76(11.52)
32.22(2.59)	15.24(11.52)
0.0(0.0)	0.04(0.28)
0.72(0.1)	0.43(0.16)
0.85(0.07)	0.4(0.29)

CV	BIC
7.0(0.0)	6.96(0.28)
15.1(4.52)	19.42(10.84)
22.9(4.52)	18.58(10.84)
0.0(0.0)	0.04(0.28)
0.49(0.08)	0.47(0.17)
0.6(0.12)	0.48(0.28)

Boxplots&tables of support recovery with p=30,N=500 and sparsity with Bernoulli(0.1)

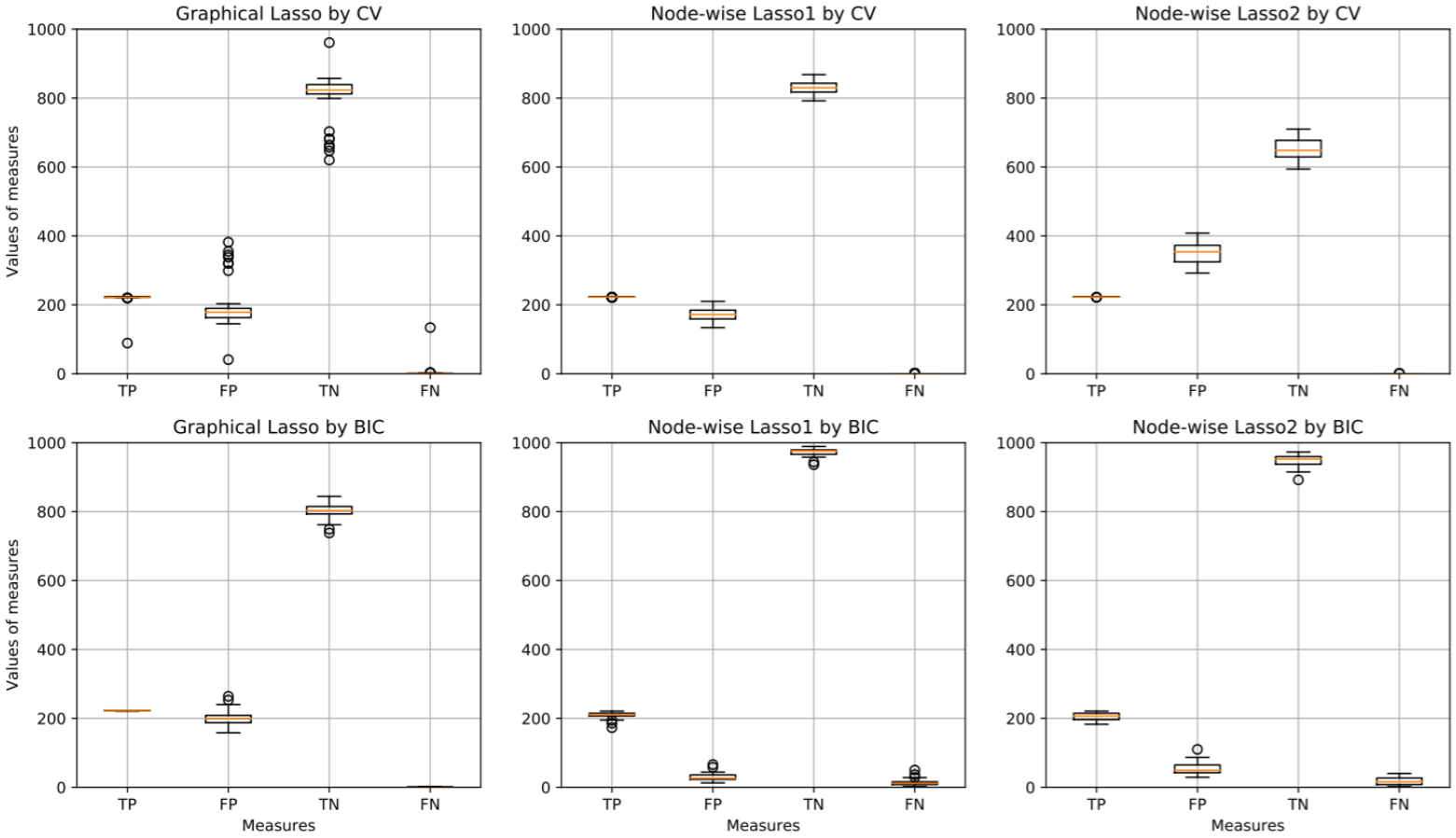


	CV	BIC
TP	53.02(21.59)	80.0(0.0)
FP	25.36(8.36)	81.94(11.17)
TN	329.64(8.36)	273.06(11.17)
FN	26.98(21.59)	0.0(0.0)
F1	0.63(0.2)	0.66(0.03)
Area	0.61(0.24)	0.77(0.03)

CV	BIC
80.0(0.0)	72.94(4.76)
55.48(9.79)	2.6(1.51)
299.52(9.79)	352.4(1.51)
0.0(0.0)	7.06(4.76)
0.74(0.03)	0.94(0.03)
0.84(0.03)	0.9(0.06)

CV	BIC
80.0(0.0)	73.74(5.38)
125.34(16.49)	18.7(9.62)
229.66(16.49)	336.3(9.62)
0.0(0.0)	6.26(5.38)
0.56(0.03)	0.86(0.04)
0.65(0.05)	0.87(0.05)

Boxplots&tables of support recovery with p=50,N=500 and sparsity with Bernoulli(0.1)

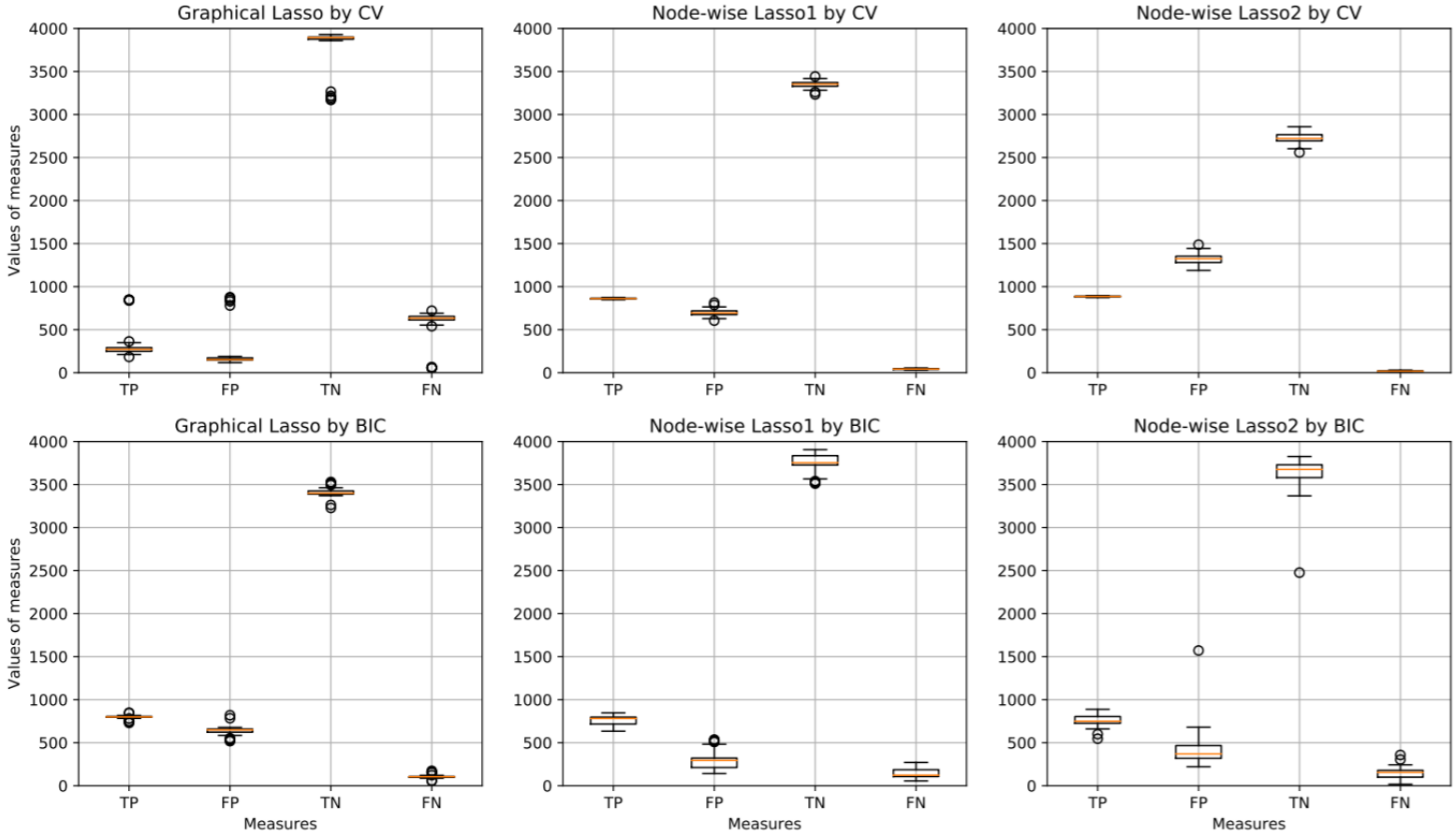


	CV	BIC
TP	219.38(18.65)	222.52(0.7)
FP	194.72(62.48)	200.52(18.95)
TN	807.28(62.48)	801.48(18.95)
FN	3.62(18.65)	0.48(0.7)
F1	0.69(0.06)	0.69(0.02)
Area	0.79(0.08)	0.8(0.02)

CV	BIC
222.78(0.46)	209.62(9.1)
171.24(17.84)	29.26(10.33)
830.76(17.84)	972.74(10.33)
0.22(0.46)	13.38(9.1)
0.72(0.02)	0.91(0.02)
0.83(0.02)	0.91(0.03)

CV	BIC
222.94(0.24)	205.52(10.72)
349.74(27.25)	54.36(16.32)
652.26(27.25)	947.64(16.32)
0.06(0.24)	17.48(10.72)
0.56(0.02)	0.85(0.02)
0.65(0.03)	0.87(0.04)

Boxplots&tables of support recovery with p=100,N=500 and sparsity with Bernoulli(0.1)

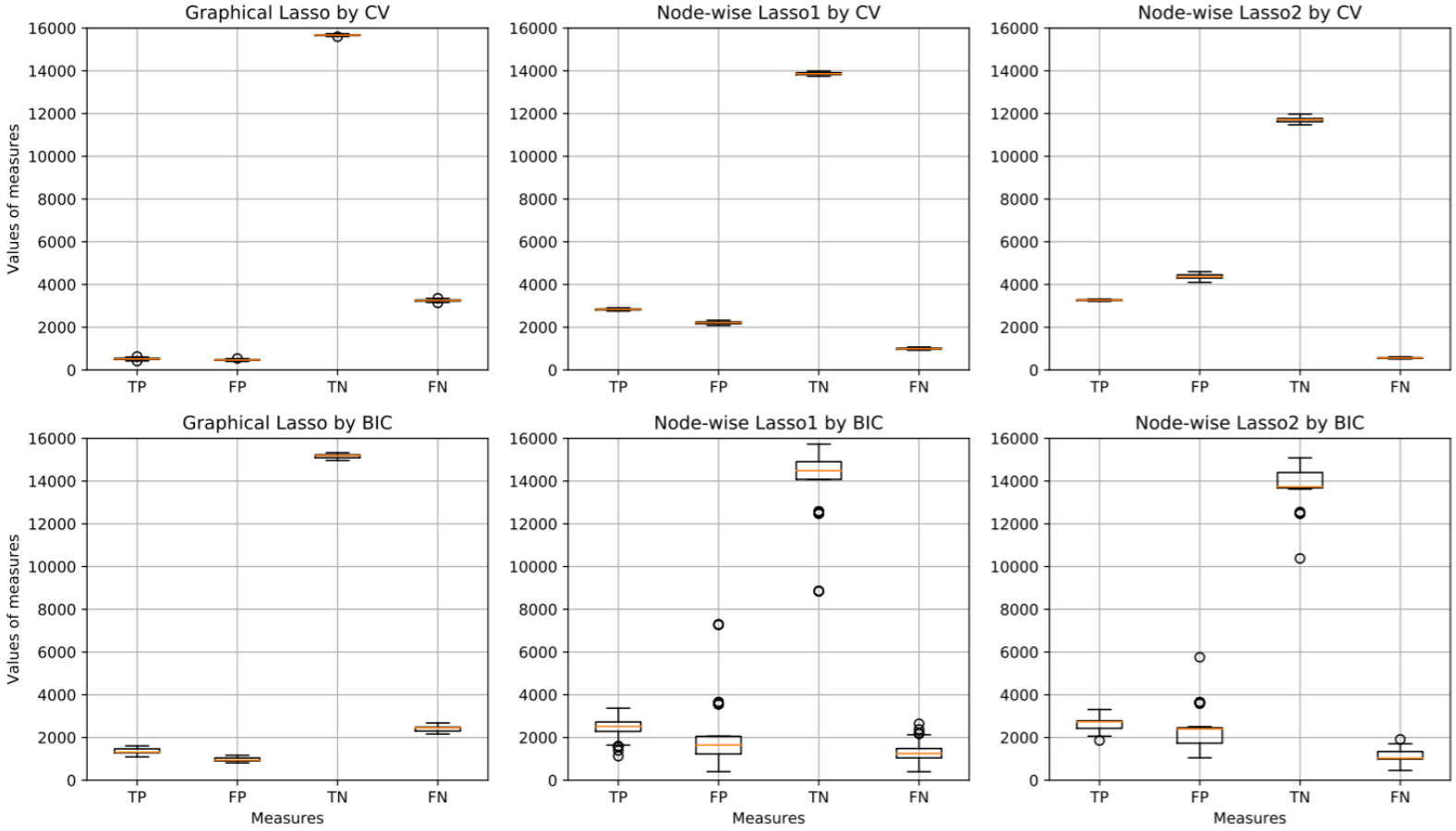


	CV	BIC
TP	325.26(176.1)	796.26(21.1)
FP	221.36(207.31)	639.98(51.29)
TN	3825.64(207.31)	3407.02(51.29)
FN	577.74(176.1)	106.74(21.1)
F1	0.43(0.08)	0.68(0.01)
Area	0.33(0.14)	0.74(0.01)

CV	BIC
860.88(6.18)	756.22(60.67)
696.44(40.24)	292.58(113.82)
3350.56(40.24)	3754.42(113.82)
42.12(6.18)	146.78(60.67)
0.7(0.01)	0.78(0.02)
0.79(0.01)	0.78(0.04)

CV	BIC
885.22(5.29)	756.12(68.86)
1326.1(62.27)	430.66(203.16)
2720.9(62.27)	3616.34(203.16)
17.78(5.29)	146.88(68.86)
0.57(0.01)	0.73(0.03)
0.66(0.01)	0.75(0.05)

Boxplots&tables of support recovery with p=200,N=500 and sparsity with Bernoulli(0.1)

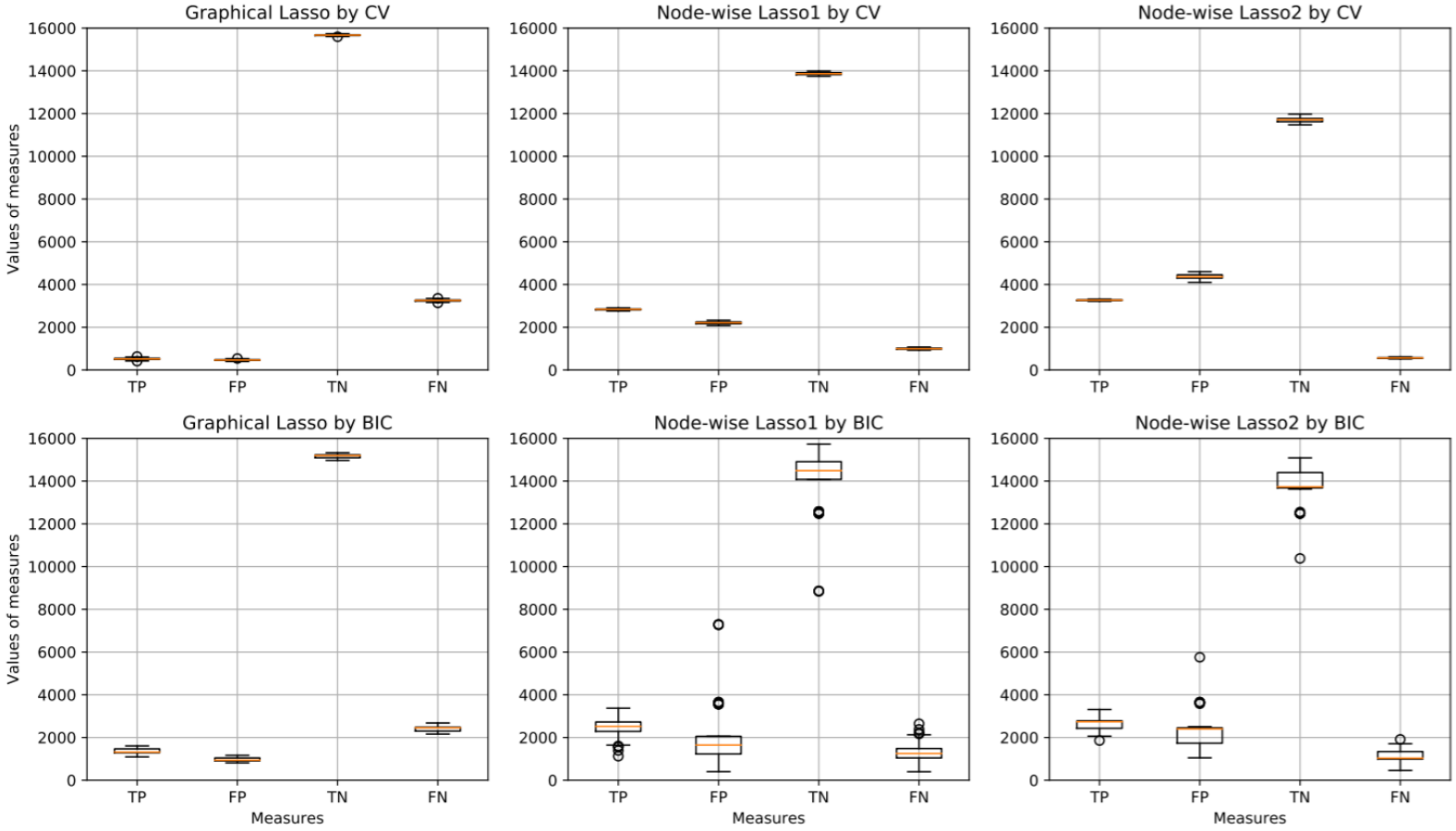


	CV	BIC
TP	517.02(46.46)	1359.72(111.14)
FP	467.06(29.81)	962.84(81.46)
TN	15664.94(29.81)	15169.16(81.46)
FN	3250.98(46.46)	2408.28(111.14)
F1	0.22(0.02)	0.45(0.02)
Area	0.13(0.01)	0.34(0.03)

CV	BIC
2834.08(33.5)	2465.06(530.24)
2208.18(66.03)	2110.46(1633.43)
13862.82(66.03)	14021.54(1633.43)
994.92(33.5)	1302.94(530.24)
0.64(0.01)	0.59(0.05)
0.64(0.01)	0.56(0.08)

CV	BIC
3267.12(23.22)	2623.92(302.74)
4362.84(109.66)	2287.24(824.86)
11708.16(109.66)	13844.76(824.86)
561.88(23.22)	1144.08(302.74)
0.57(0.0)	0.61(0.02)
0.62(0.01)	0.59(0.04)

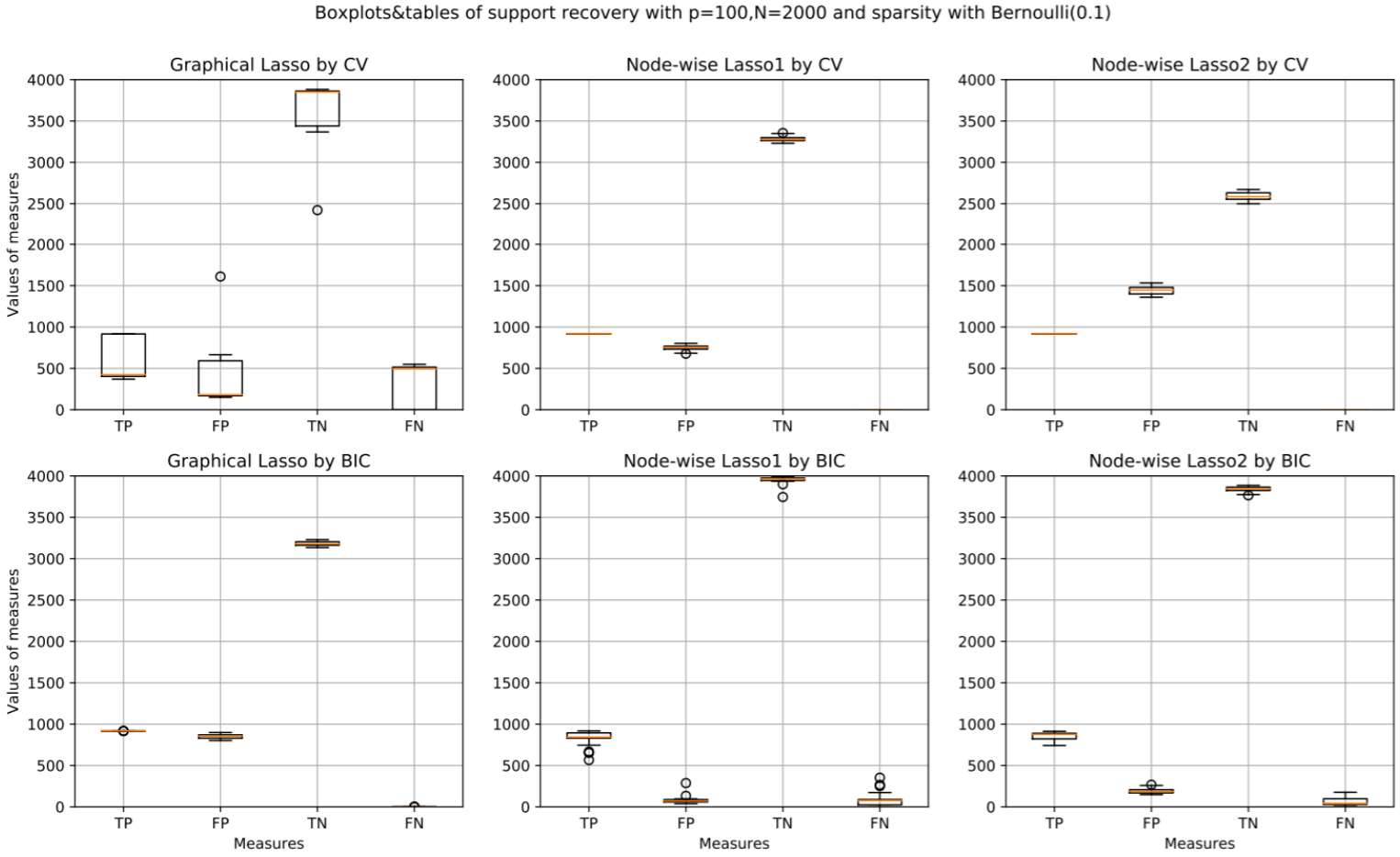
Boxplots&tables of support recovery with p=200,N=500 and sparsity with Bernoulli(0.1)



	CV	BIC
TP	517.02(46.46)	1359.72(111.14)
FP	467.06(29.81)	962.84(81.46)
TN	15664.94(29.81)	15169.16(81.46)
FN	3250.98(46.46)	2408.28(111.14)
F1	0.22(0.02)	0.45(0.02)
Area	0.13(0.01)	0.34(0.03)

CV	BIC
2834.08(33.5)	2465.06(530.24)
2208.18(66.03)	2110.46(1633.43)
13862.82(66.03)	14021.54(1633.43)
994.92(33.5)	1302.94(530.24)
0.64(0.01)	0.59(0.05)
0.64(0.01)	0.56(0.08)

CV	BIC
3267.12(23.22)	2623.92(302.74)
4362.84(109.66)	2287.24(824.86)
11708.16(109.66)	13844.76(824.86)
561.88(23.22)	1144.08(302.74)
0.57(0.0)	0.61(0.02)
0.62(0.01)	0.59(0.04)

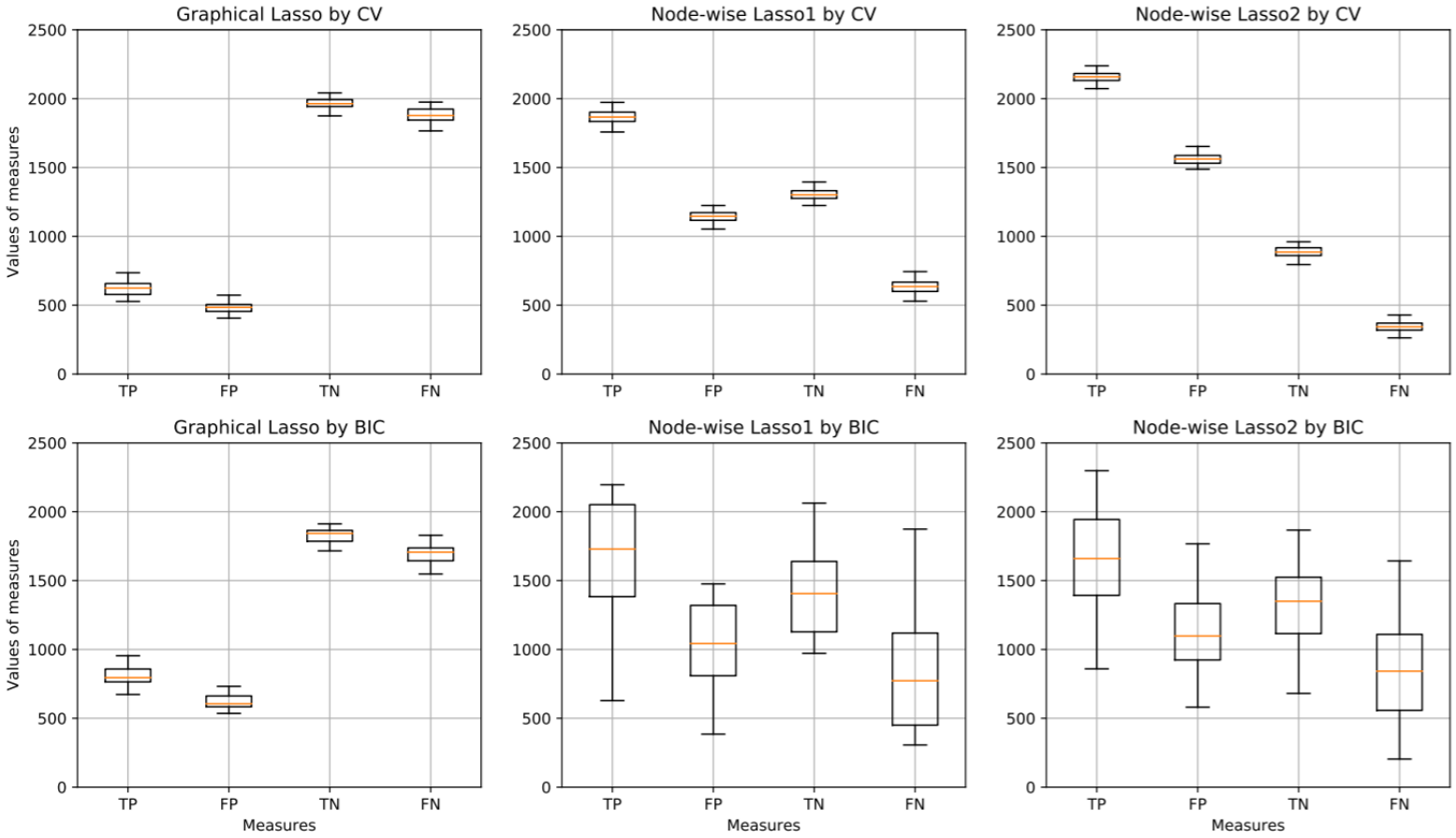


	CV	BIC
TP	600.58(247.71)	917.94(0.24)
FP	356.36(274.31)	852.28(25.18)
TN	3675.64(274.31)	3179.72(25.18)
FN	317.42(247.71)	0.06(0.24)
F1	0.62(0.1)	0.68(0.01)
Area	0.58(0.2)	0.79(0.01)

CV	BIC
918.0(0.0)	837.48(74.75)
751.54(30.27)	75.76(34.89)
3280.46(30.27)	3956.24(34.89)
0.0(0.0)	80.52(74.75)
0.71(0.01)	0.91(0.04)
0.81(0.01)	0.89(0.08)

CV	BIC
918.0(0.0)	858.64(41.54)
1441.32(43.85)	193.58(33.25)
2590.68(43.85)	3838.42(33.25)
0.0(0.0)	59.36(41.54)
0.56(0.01)	0.87(0.01)
0.64(0.01)	0.89(0.04)

Boxplots&tables of support recovery with p=100,N=2000 and sparsity with Bernoulli(0.3)

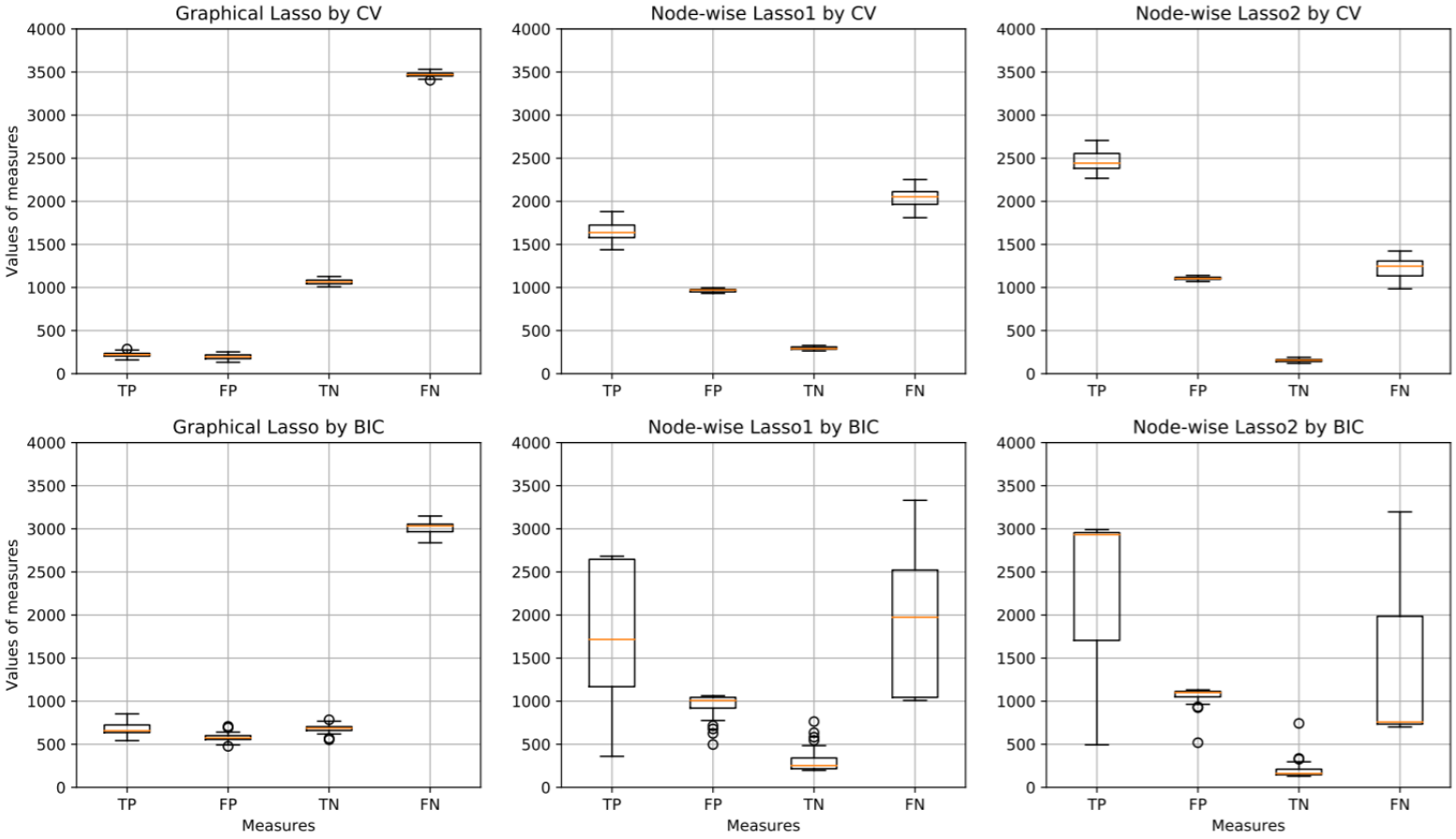


	CV	BIC
TP	623.76(49.45)	805.48(65.21)
FP	482.98(40.58)	620.34(45.84)
TN	1965.02(40.58)	1827.66(45.84)
FN	1878.24(49.45)	1696.52(65.21)
F1	0.35(0.02)	0.41(0.02)
Area	0.2(0.01)	0.24(0.01)

CV	BIC
1869.18(49.65)	1649.9(406.92)
1144.92(42.02)	1020.42(294.89)
1303.08(42.02)	1427.58(294.89)
632.82(49.65)	852.1(406.92)
0.68(0.01)	0.63(0.08)
0.4(0.01)	0.37(0.04)

CV	BIC
2157.12(36.0)	1638.64(325.66)
1562.4(42.37)	1112.22(263.95)
885.6(42.37)	1335.78(263.95)
344.88(36.0)	863.36(325.66)
0.69(0.01)	0.62(0.06)
0.31(0.01)	0.34(0.03)

Boxplots&tables of support recovery with p=100,N=2000 and sparsity with Bernoulli(0.5)



	CV	BIC
TP	219.94(26.63)	667.82(61.79)
FP	195.86(28.31)	578.84(44.61)
TN	1064.14(28.31)	681.16(44.61)
FN	3470.06(26.63)	3022.18(61.79)
F1	0.11(0.01)	0.27(0.02)
Area	0.05(0.0)	0.1(0.0)

CV	BIC
1644.84(93.89)	1924.12(760.54)
965.94(17.47)	962.66(123.46)
294.06(17.47)	297.34(123.46)
2045.16(93.89)	1765.88(760.54)
0.52(0.02)	0.56(0.17)
0.1(0.01)	0.11(0.02)

CV	BIC
2469.08(106.56)	2382.3(686.75)
1104.96(16.54)	1066.8(94.3)
155.04(16.54)	193.2(94.3)
1220.92(106.56)	1307.7(686.75)
0.68(0.02)	0.65(0.14)
0.08(0.01)	0.09(0.01)