Collinearity in Prognostic Models for Dysphagia

Artuur Leeuwenberg, PhD

Julius Center - University Medical Center Utrecht
With Ewoud Schuit, Hans Reitsma, and Carl Moons

Compared methods

1. Standard logistic regression (**LR**)
2. **Lasso** penalization
3. **Ridge** penalization
4. **ElasticNet** penalization
5. Principal component analysis logistic regression (**PCLR**)
6. Linear auto-encoder logistic regression (**LAELR**)
7. **Dropout** regularization
8. Non-negative constraints (**LRnn**)

Simulation study

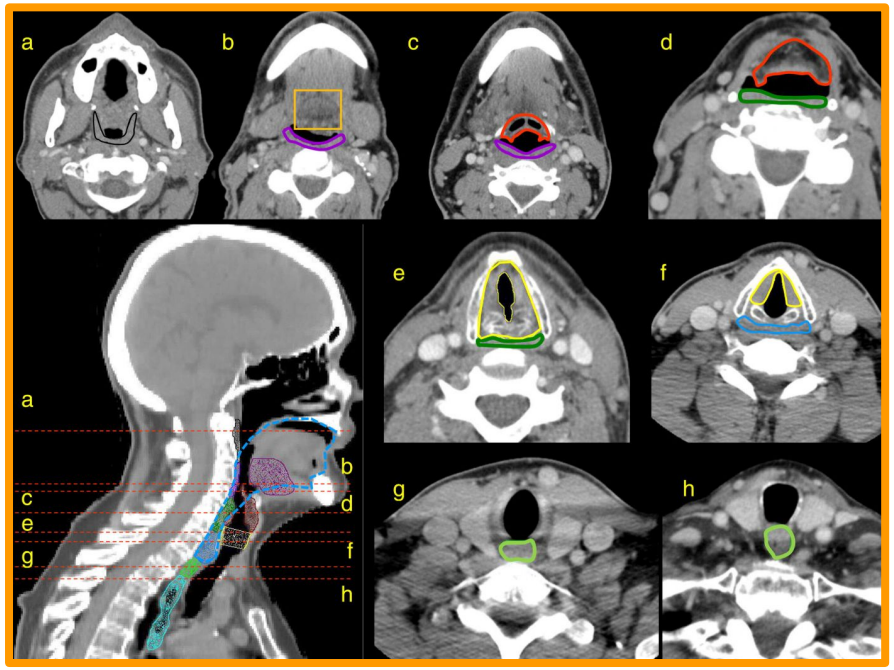
Case study (simulation study):
For **head and neck cancer patients**, we aim to predict the risk of experiencing **complications** 6 months **after radiotherapy**.

Predictors X :

- **Dosage per relevant organ**
- …

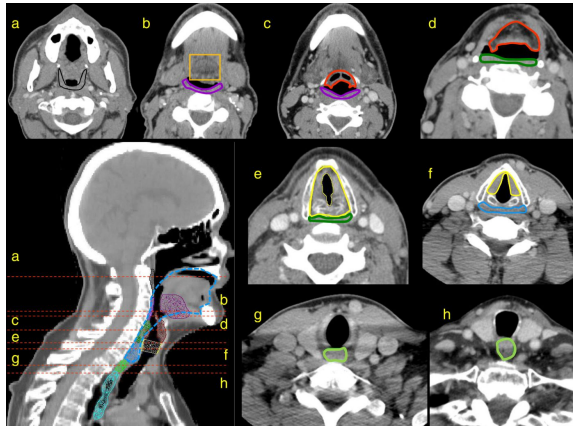
Outcome y:

- Dysphagia M_6



High **collinearity** of dosage for OAR

	AGE	Subm.L.Dm	Subm.R.Dm	Parotid.L.Dm	Parotid.R.Dm	PCM.Sup.Dm	PCM.Med.Dm	PCM.Inf.Dm	Supraglottic.Dm	OralCavity.Ext.Dm	GlotticArea.Dm	DYSFAGIE.BSL.1
Subm.L.Dm	-0.25											
Subm.R.Dm	-0.27	0.88										
Parotid.L.Dm	-0.26	0.85	0.72									
Parotid.R.Dm	-0.28	0.69	0.83	0.74								
PCM.Sup.Dm	-0.28	0.85	0.86	0.82	0.81							
PCM.Med.Dm	-0.25	0.87	0.89	0.74	0.72	0.79						
PCM.Inf.Dm	0.11	-0.11	-0.13	-0.11	-0.12	-0.35	0.11					
Supraglottic.Dm	-0.08	0.36	0.38	0.27	0.27	0.12	0.60	0.69				
OralCavity.Ext.Dm	-0.28	0.83	0.85	0.80	0.80	0.95	0.73	-0.40	0.07			
GlotticArea.Dm	0.15	-0.33	-0.34	-0.31	-0.31	-0.56	-0.16	0.89	0.54	-0.60		
DYSFAGIE.BSL.1	-0.04	0.24	0.25	0.25	0.21	0.24	0.23	-0.06	0.03	0.28	-0.12	
DYSFAGIE.BSL.2	-0.09	0.18	0.17	0.23	0.23	0.17	0.20	0.06	0.13	0.21	-0.01	-0.10



Simulation setup

Two levels of collinearity: **2** settings.

- Multivariate Gaussian to simulate predictor (cov from real data of 750 pts.)
- True predictor–outcome relation: Ridge fit on real data of 750 pts.
- ~600 training samples, 10.000 test samples, 100 repeats

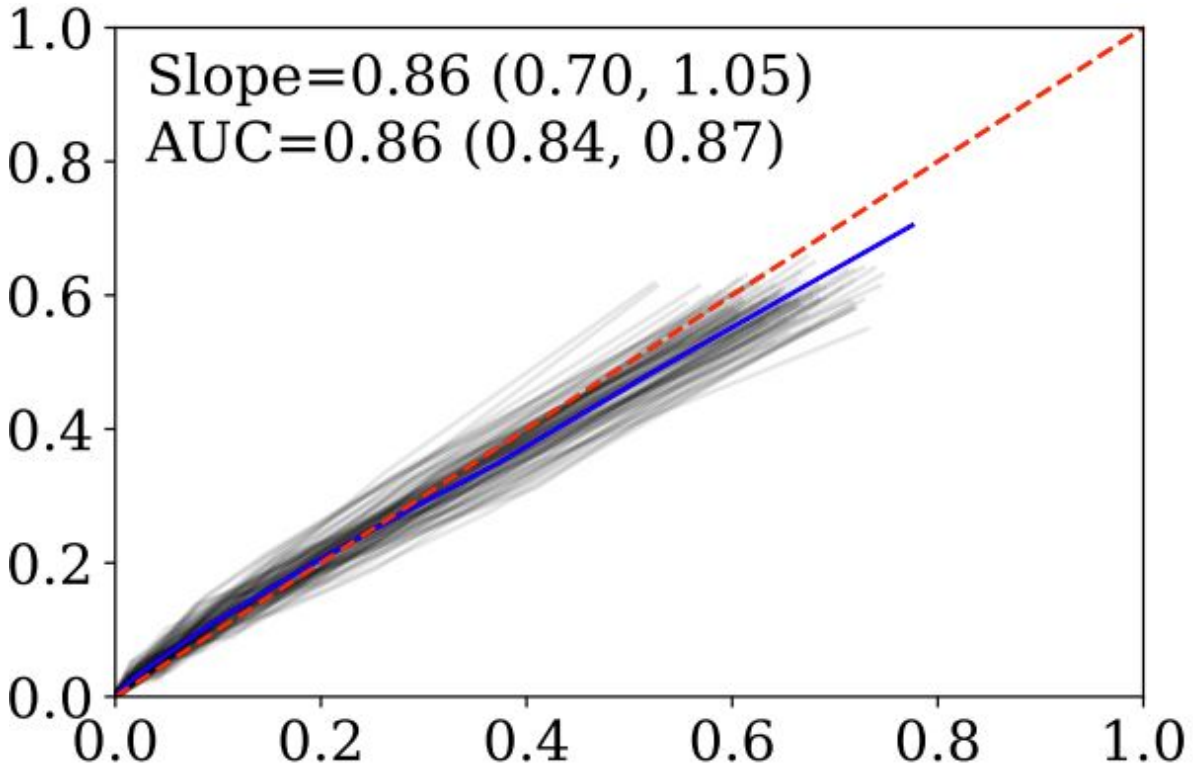
* We only discuss a selection of the studied settings.

Results: spot the differences

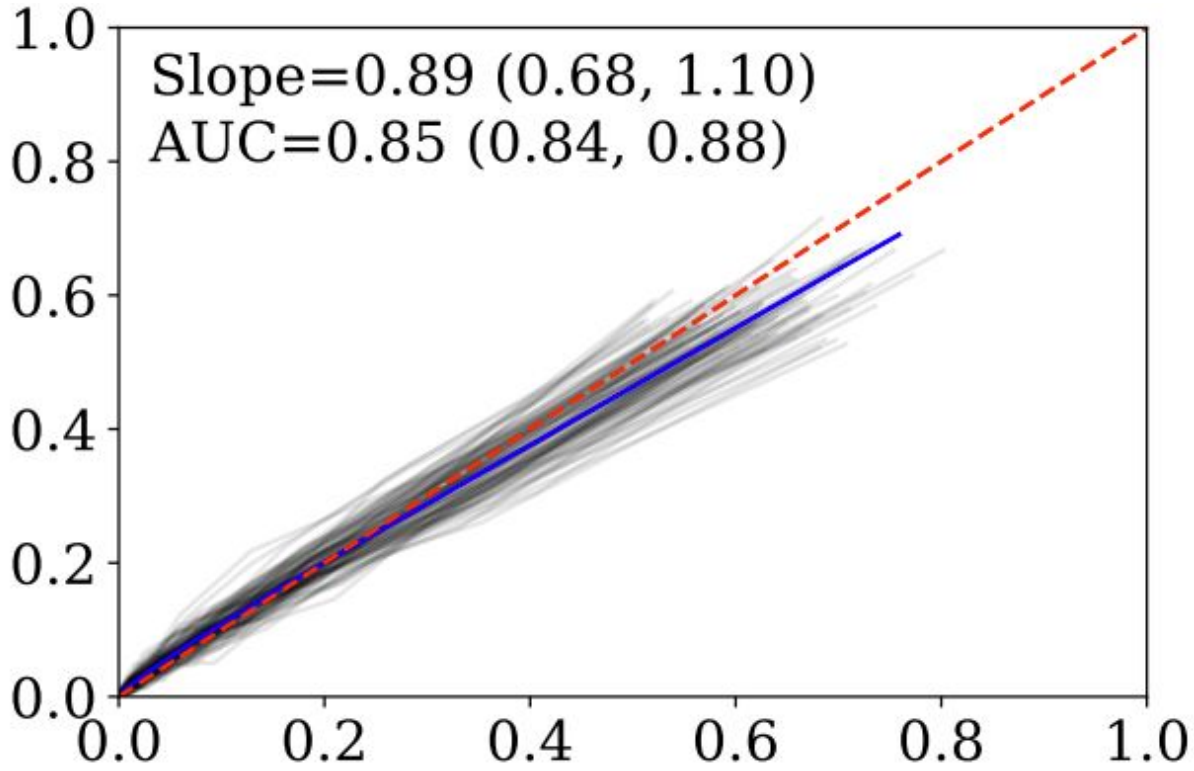
VIF=5 (EPV:6)

VIF=43 (EPV:6)

Results: LR

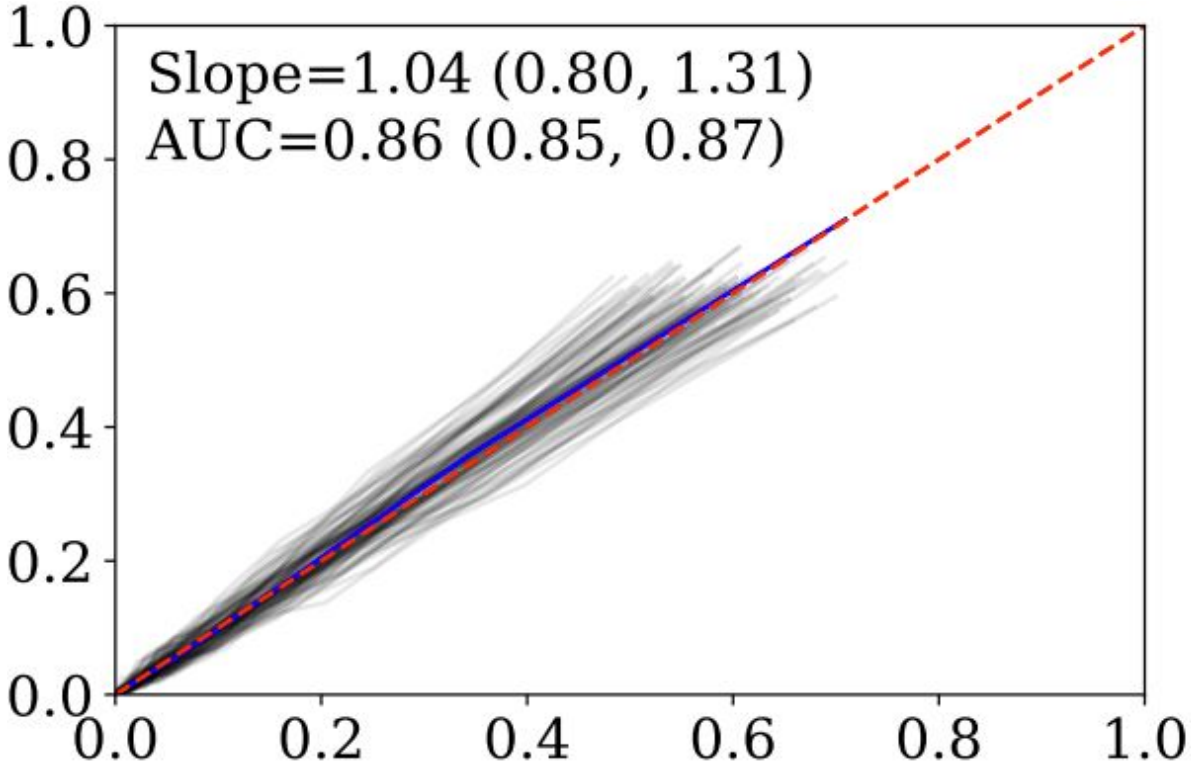


VIF=5 (EPV:6)

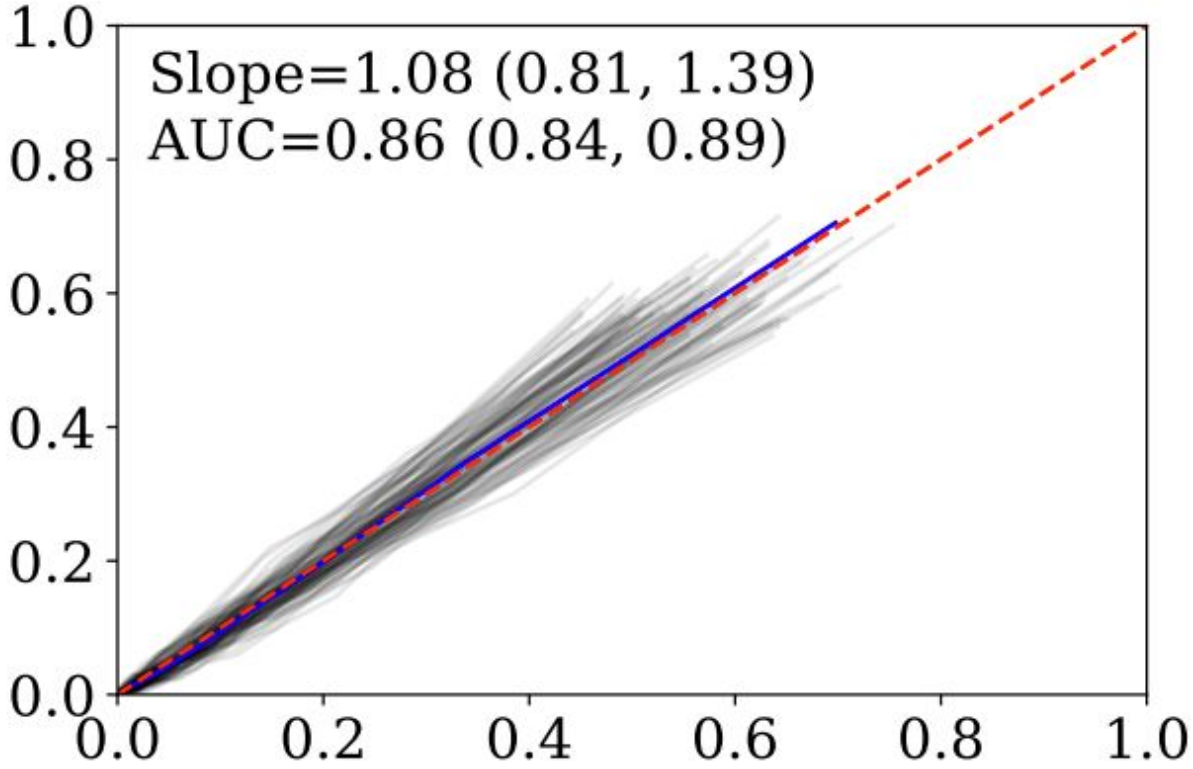


VIF=43 (EPV:6)

Results: Lasso

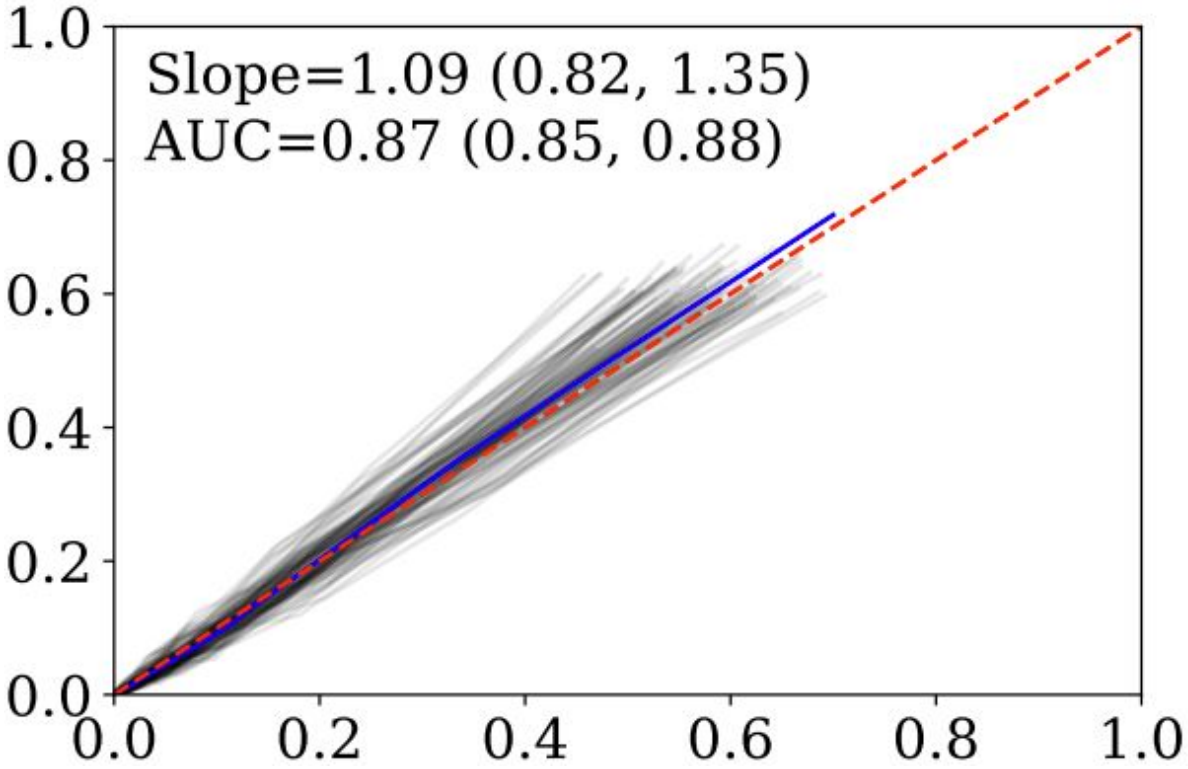


VIF=5 (EPV:6)

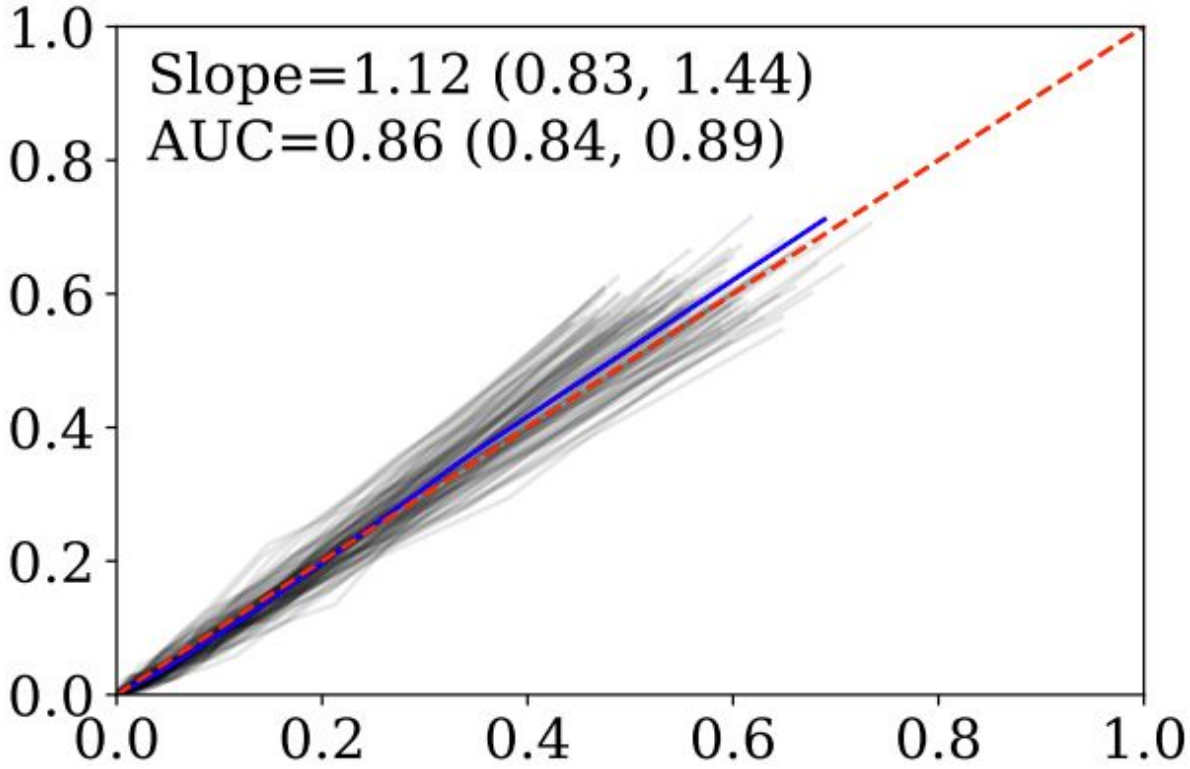


VIF=43 (EPV:6)

Results: Ridge

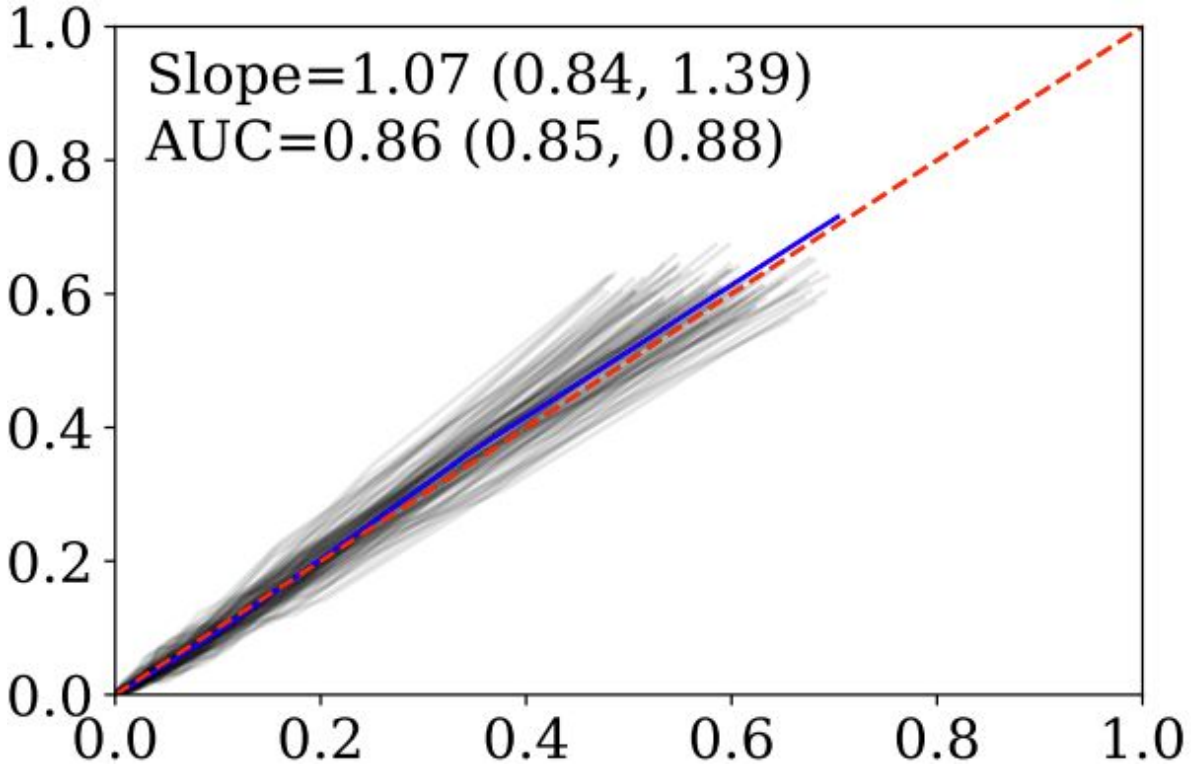


VIF=5 (EPV:6)

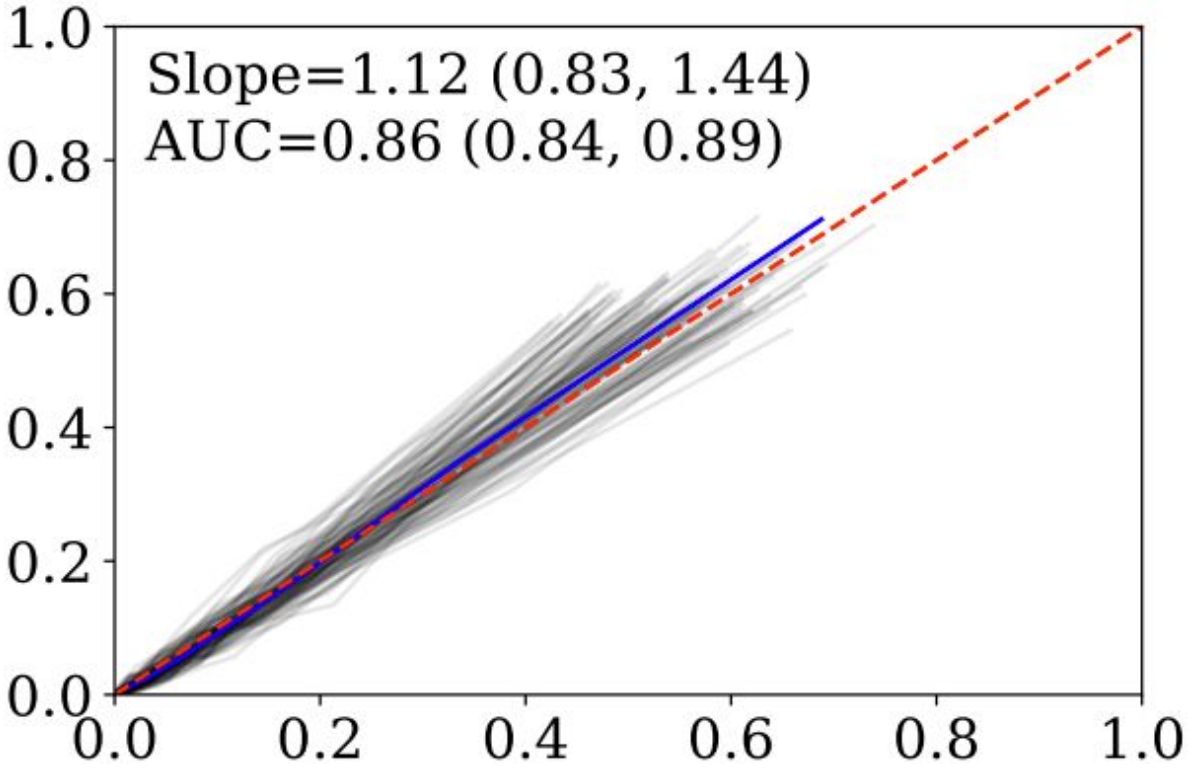


VIF=43 (EPV:6)

Results: ElasticNet

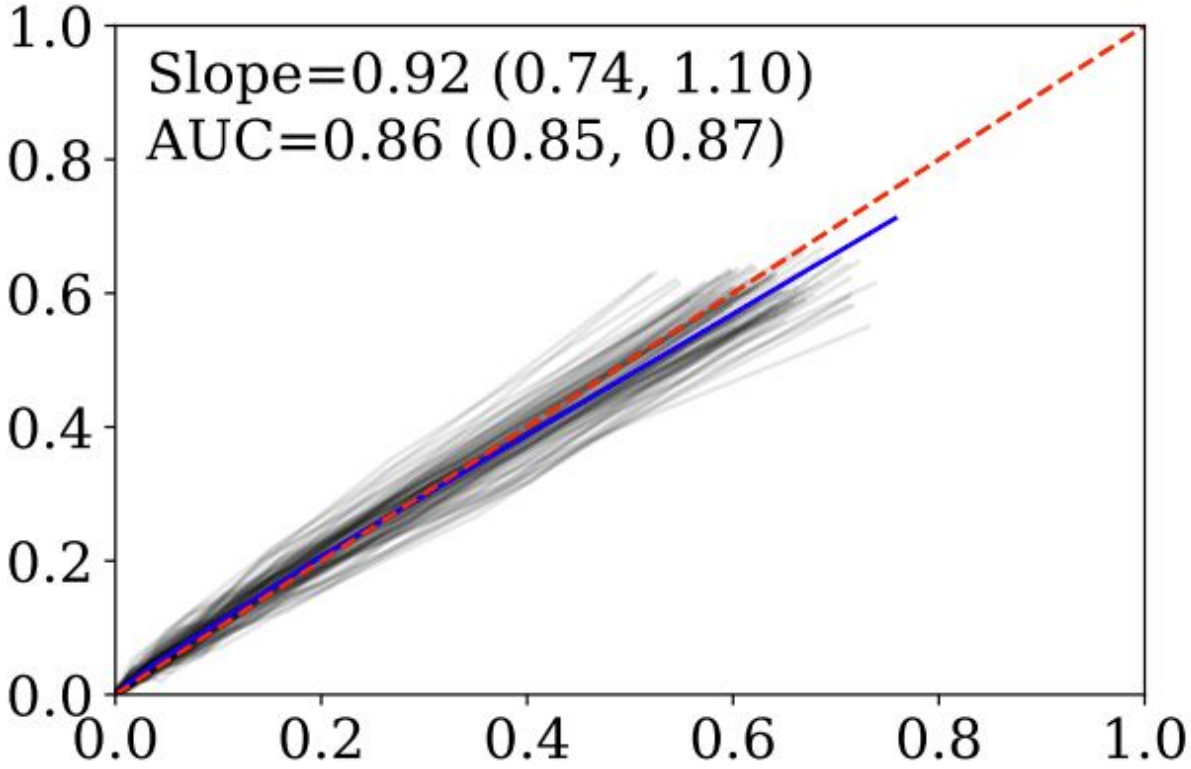


VIF=5 (EPV:6)

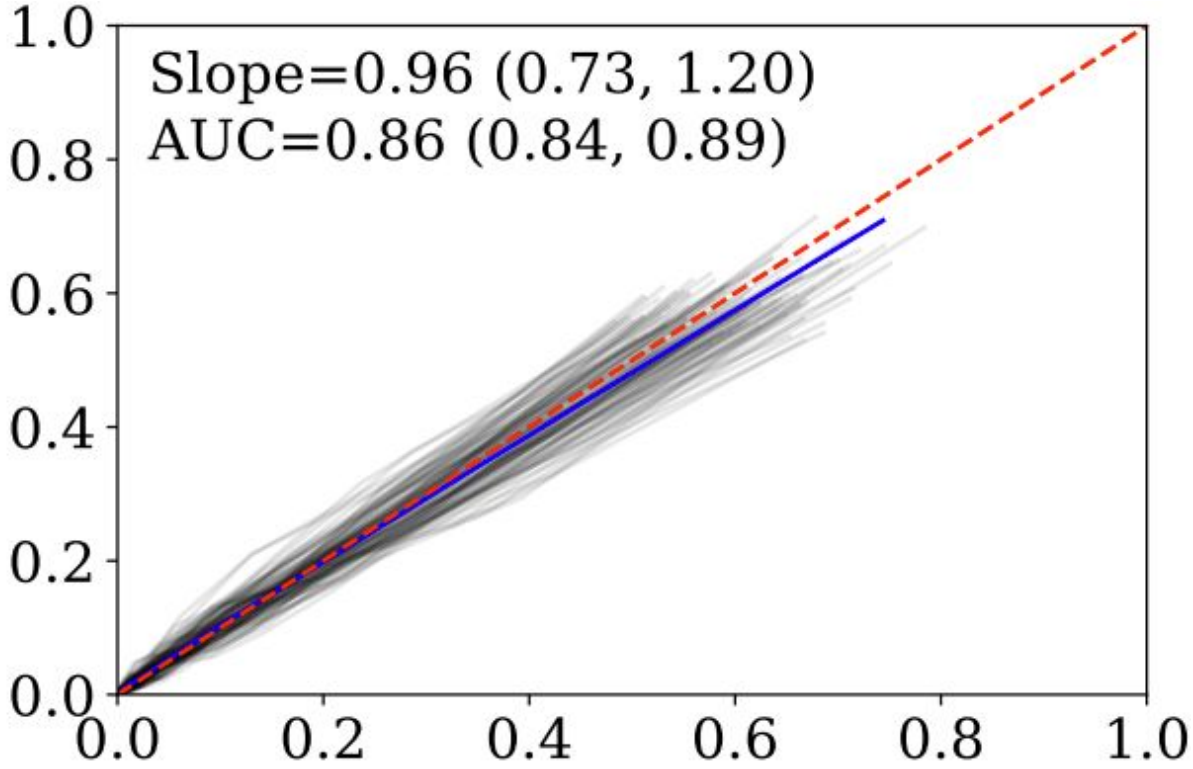


VIF=43 (EPV:6)

Results: PCLR

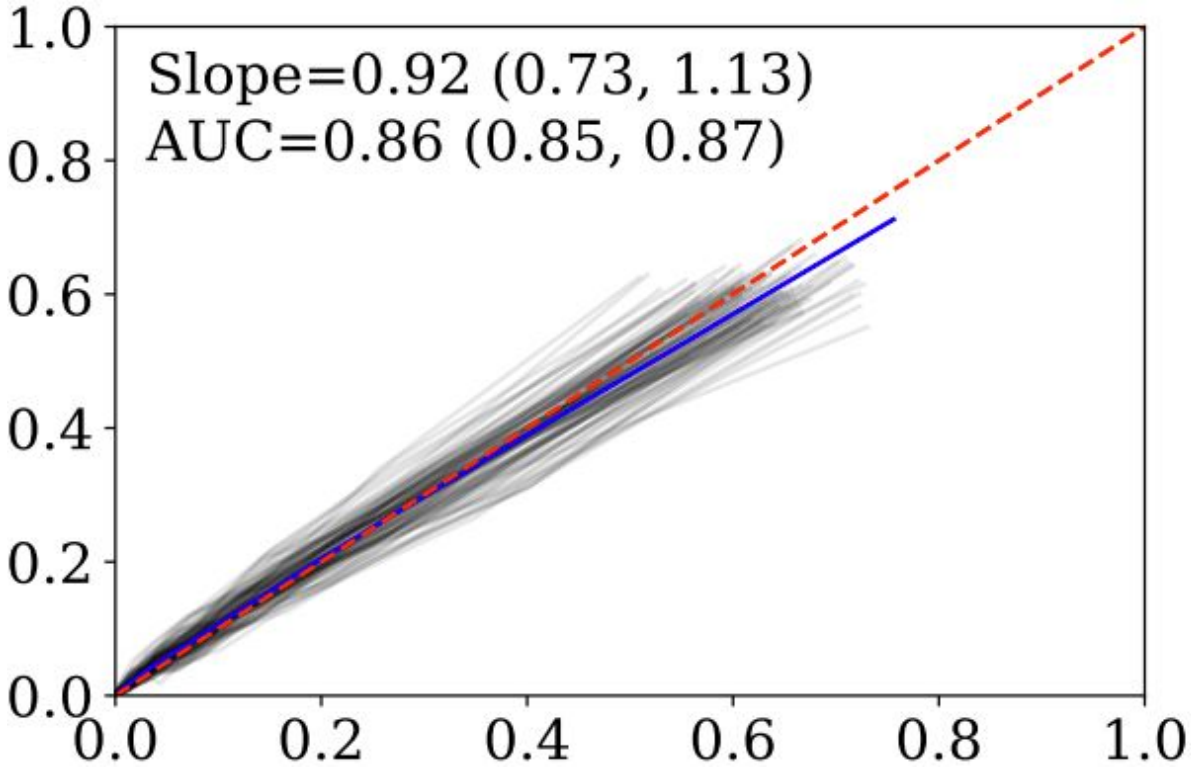


VIF=5 (EPV:6)

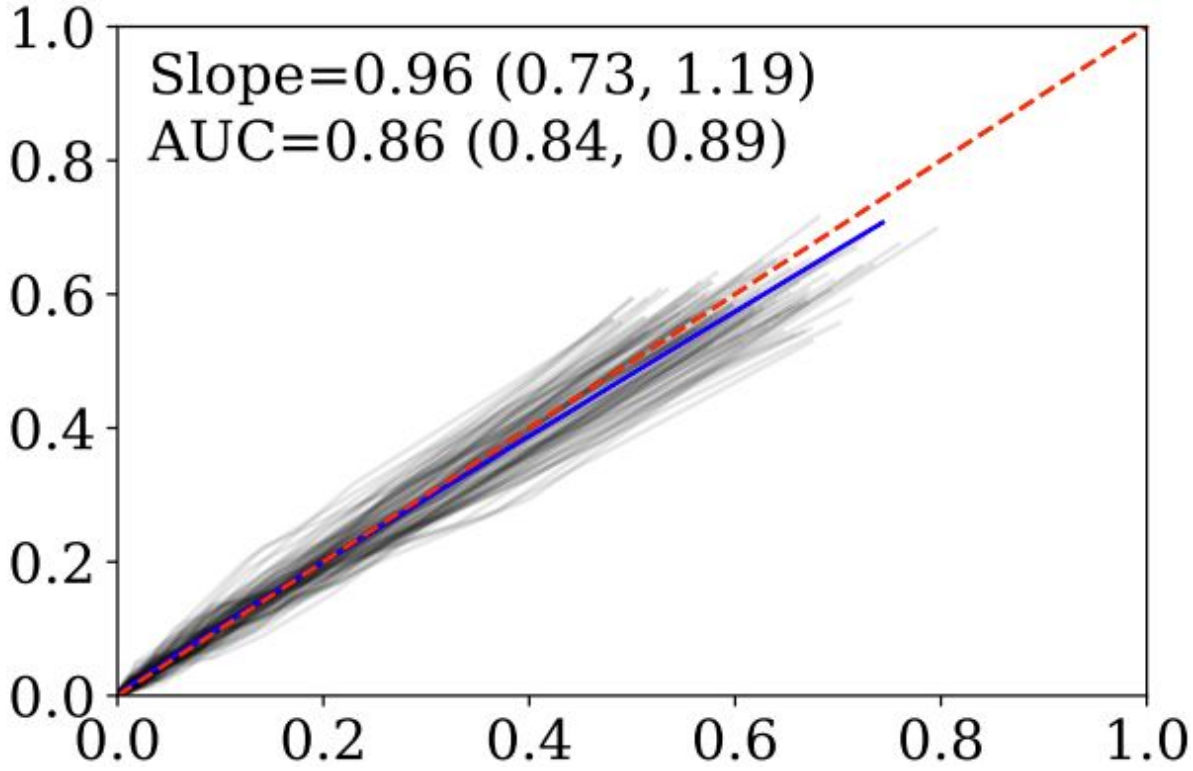


VIF=43 (EPV:6)

Results: LAELR

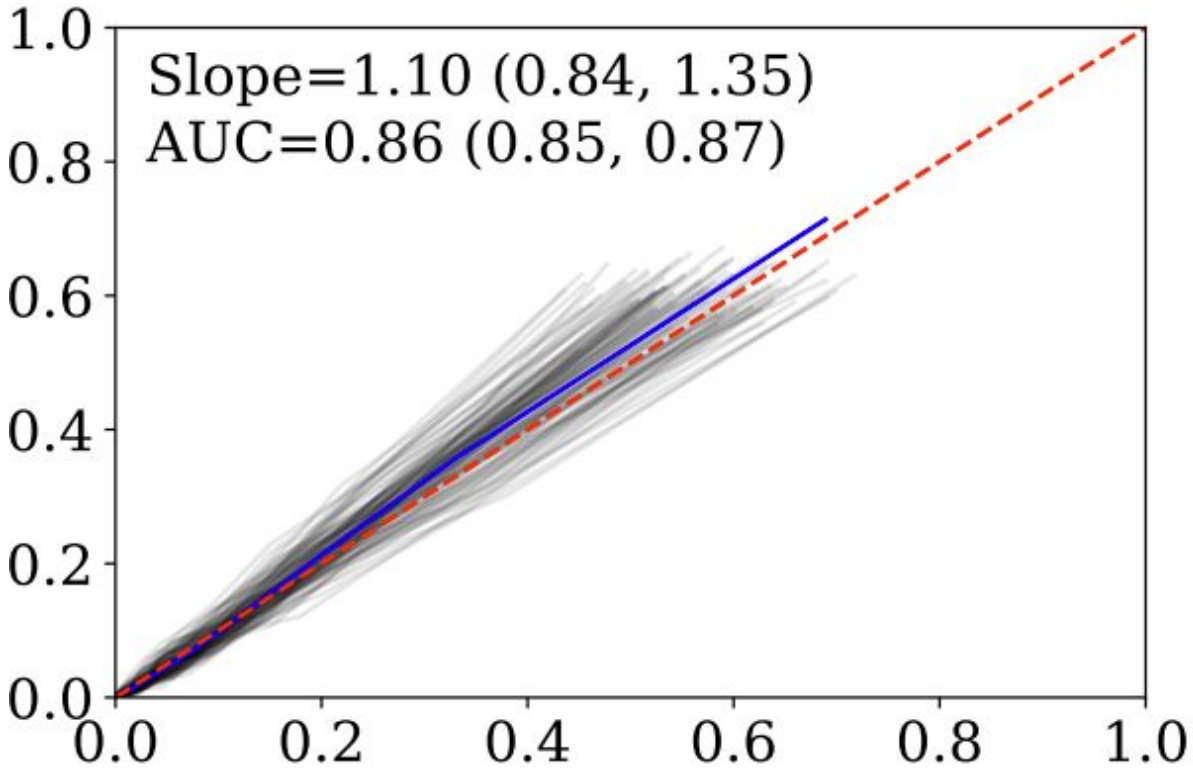


VIF=5 (EPV:6)

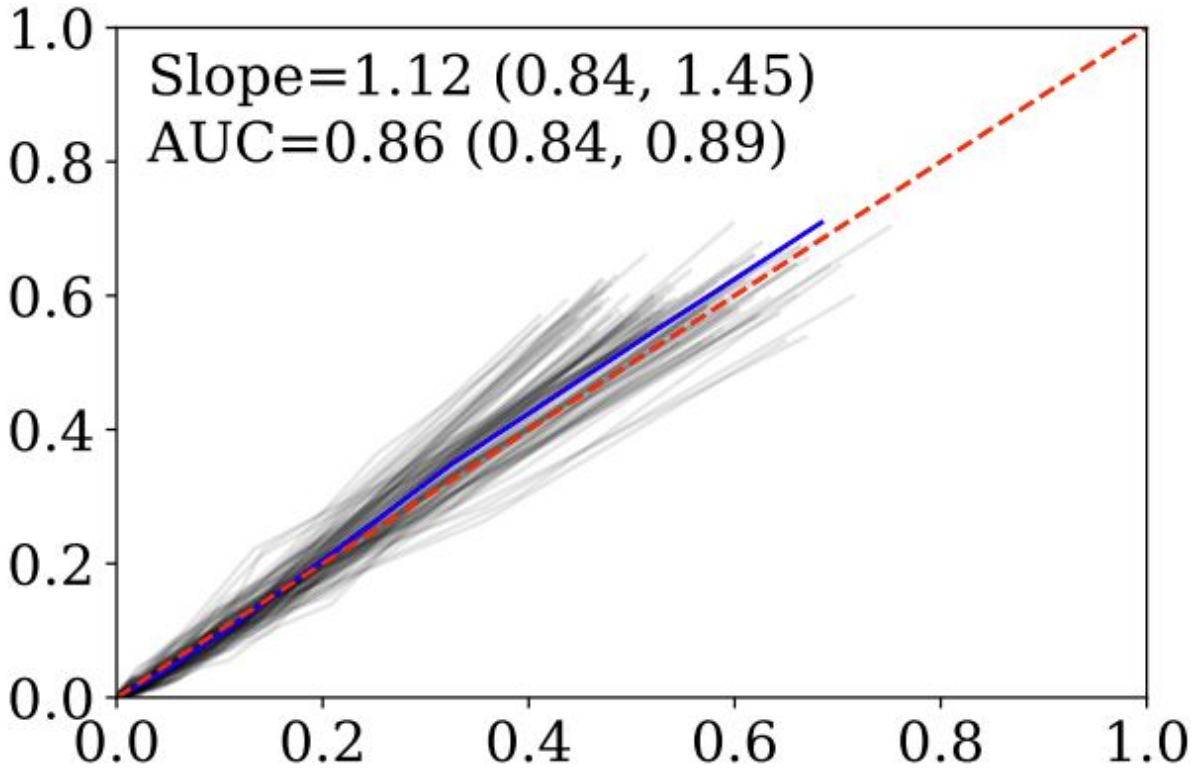


VIF=43 (EPV:6)

Results: Dropout

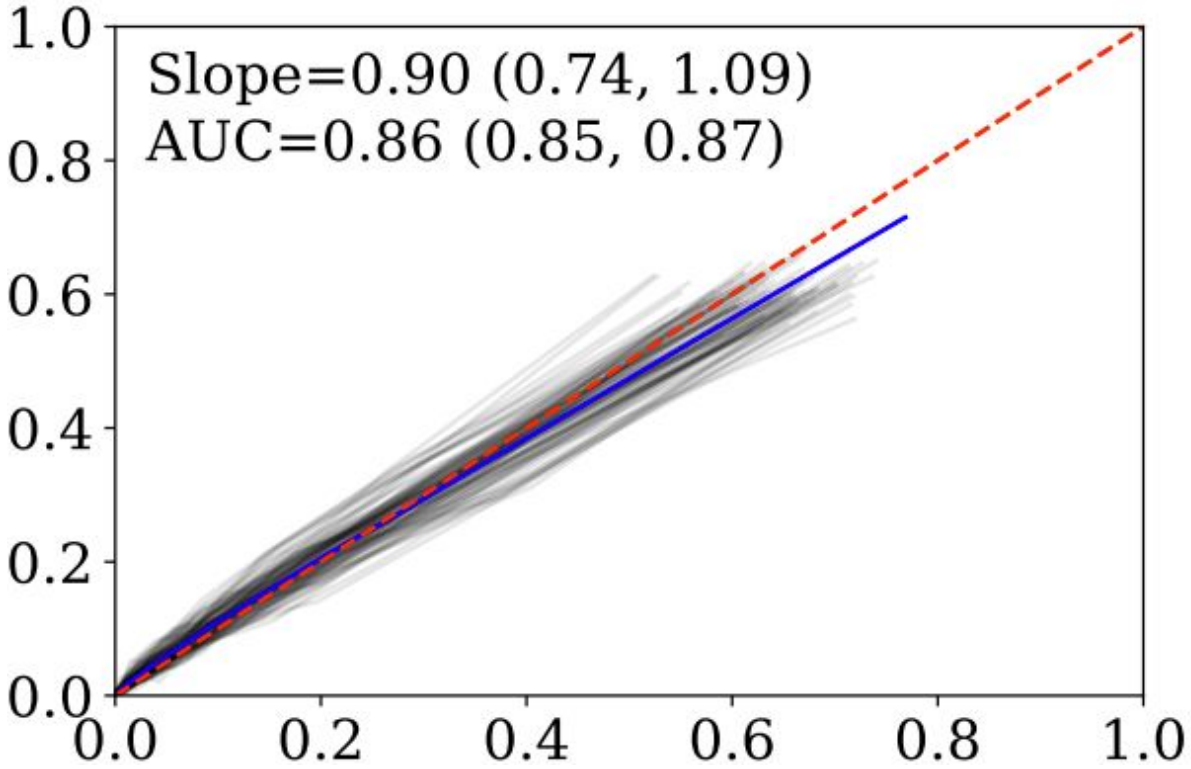


VIF=5 (EPV:6)

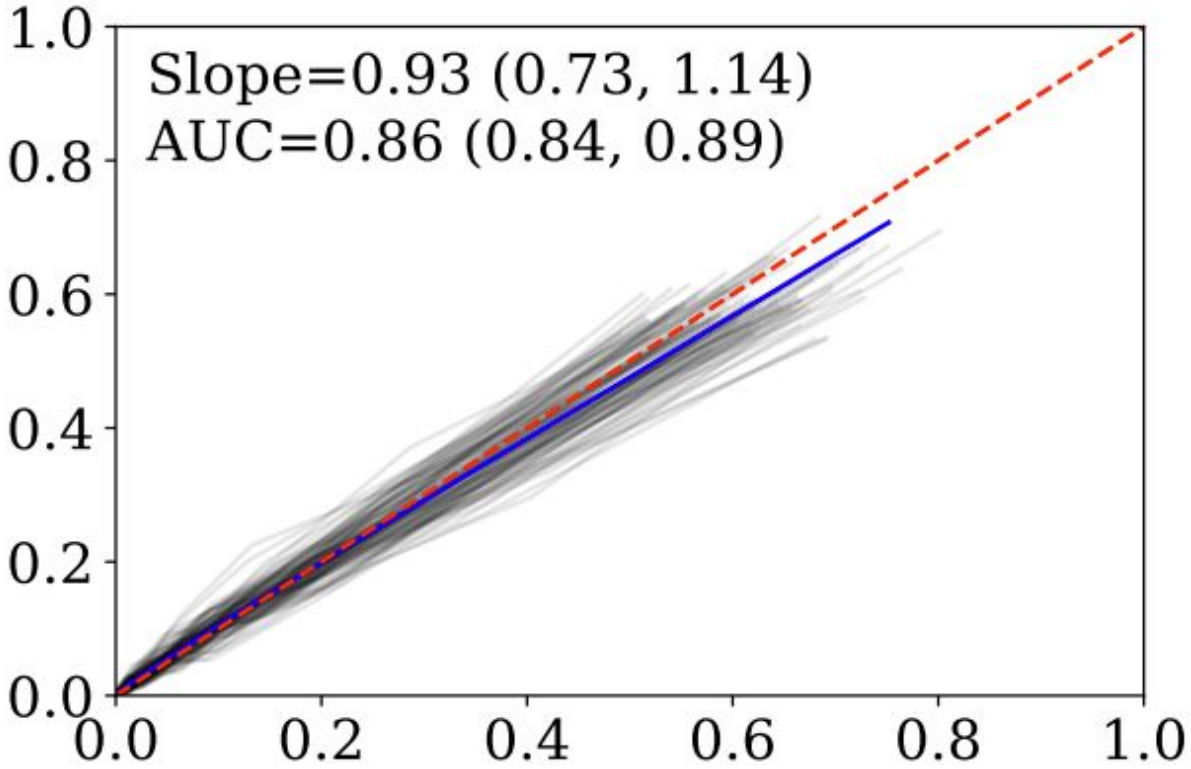


VIF=43 (EPV:6)

Results: LR_{NN}



VIF=5 (EPV:6)

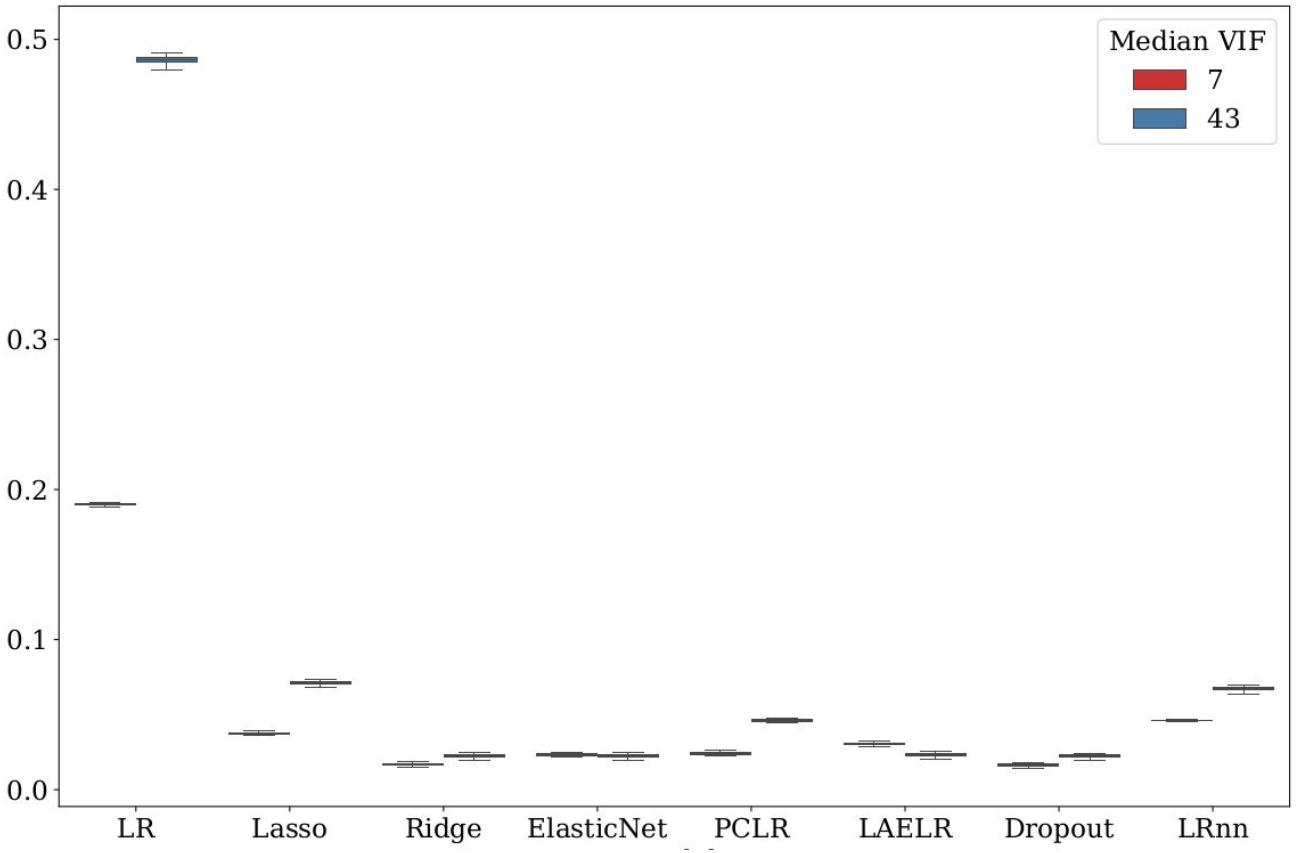


VIF=43 (EPV:6)

Confirmation

Collinearity did not impact predictive performance in the simulations.

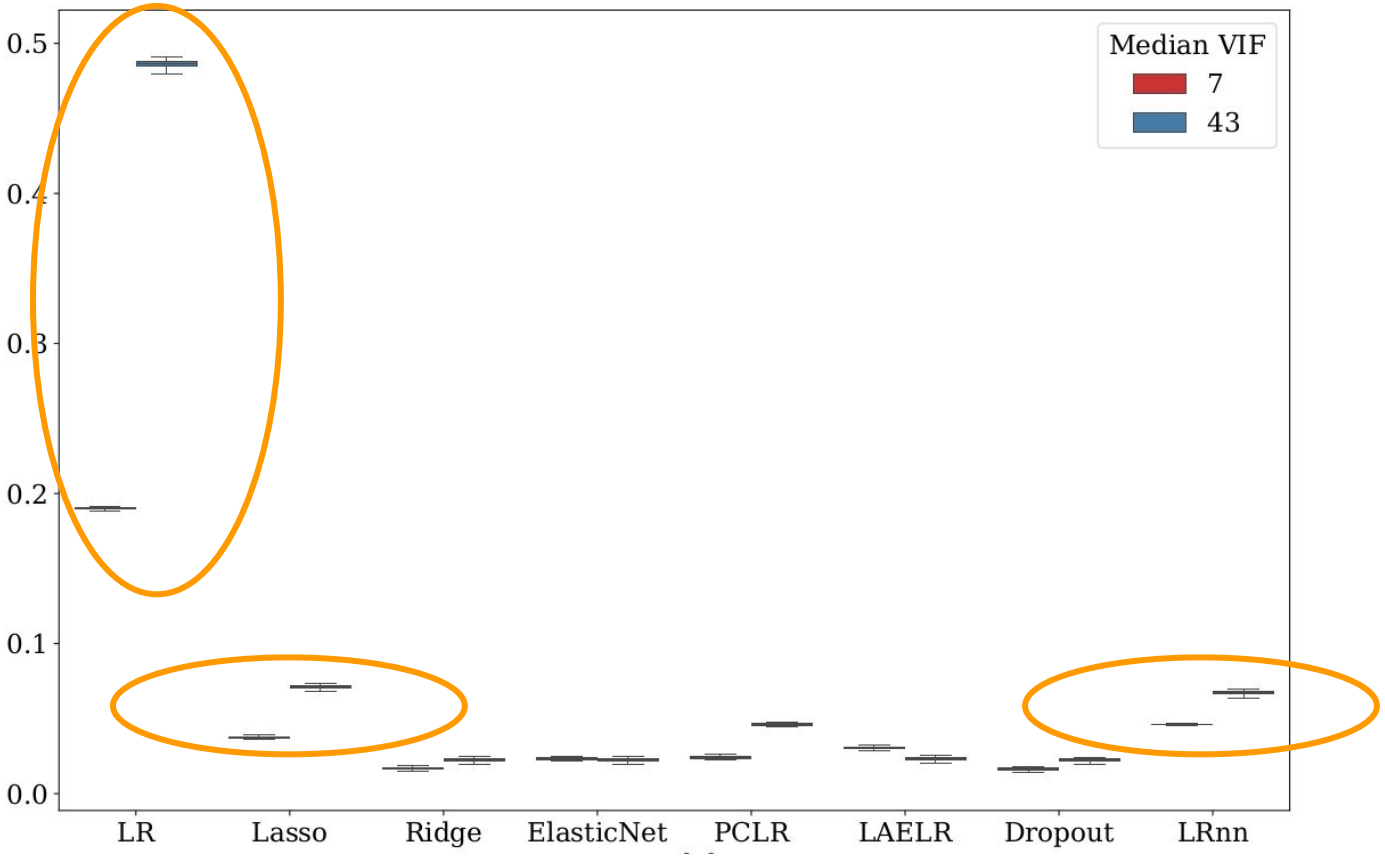
Coefficient estimation



MSE

How far are the estimated coefficients from the true coefficients?

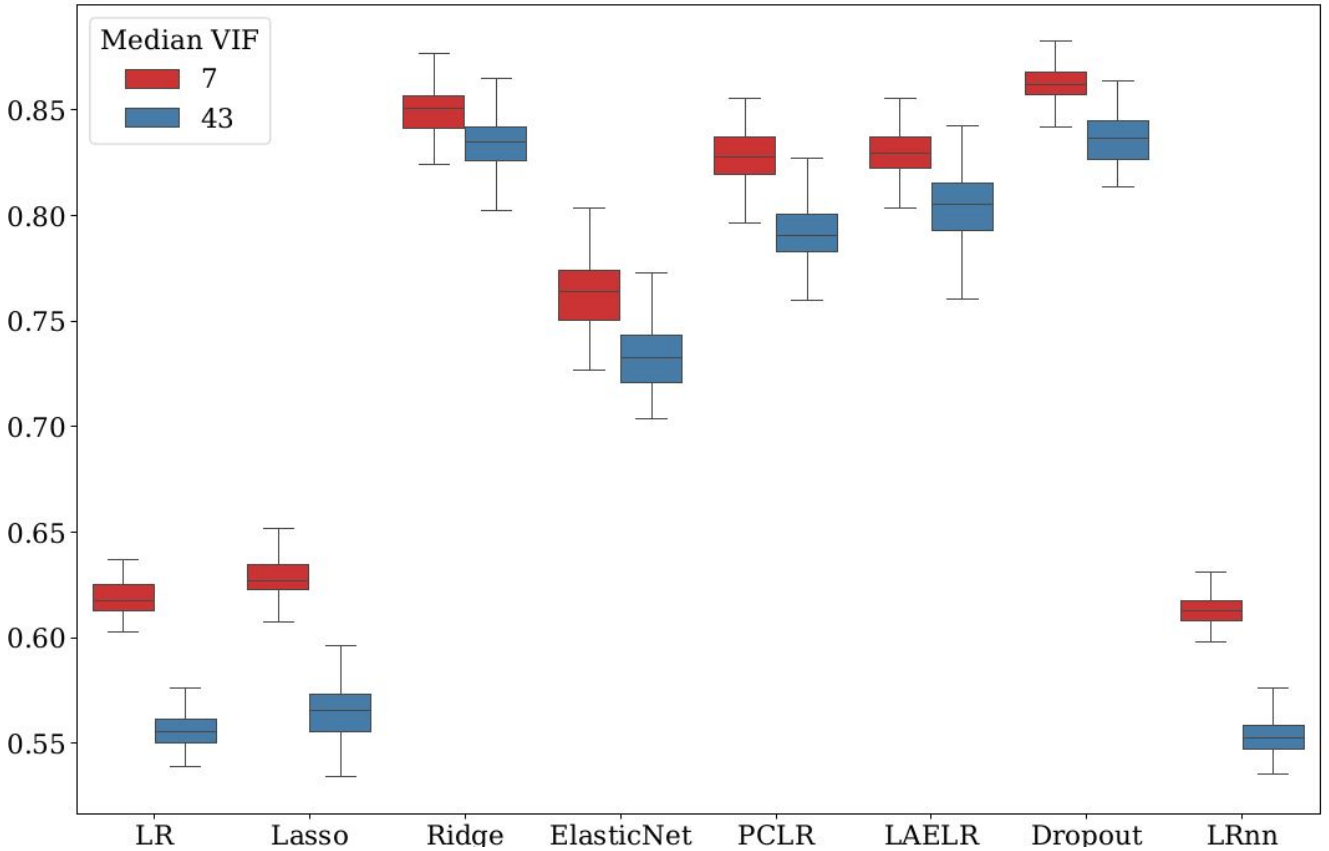
Coefficient estimation



MSE

How far are the estimated coefficients from the true coefficients?

Coefficient estimation

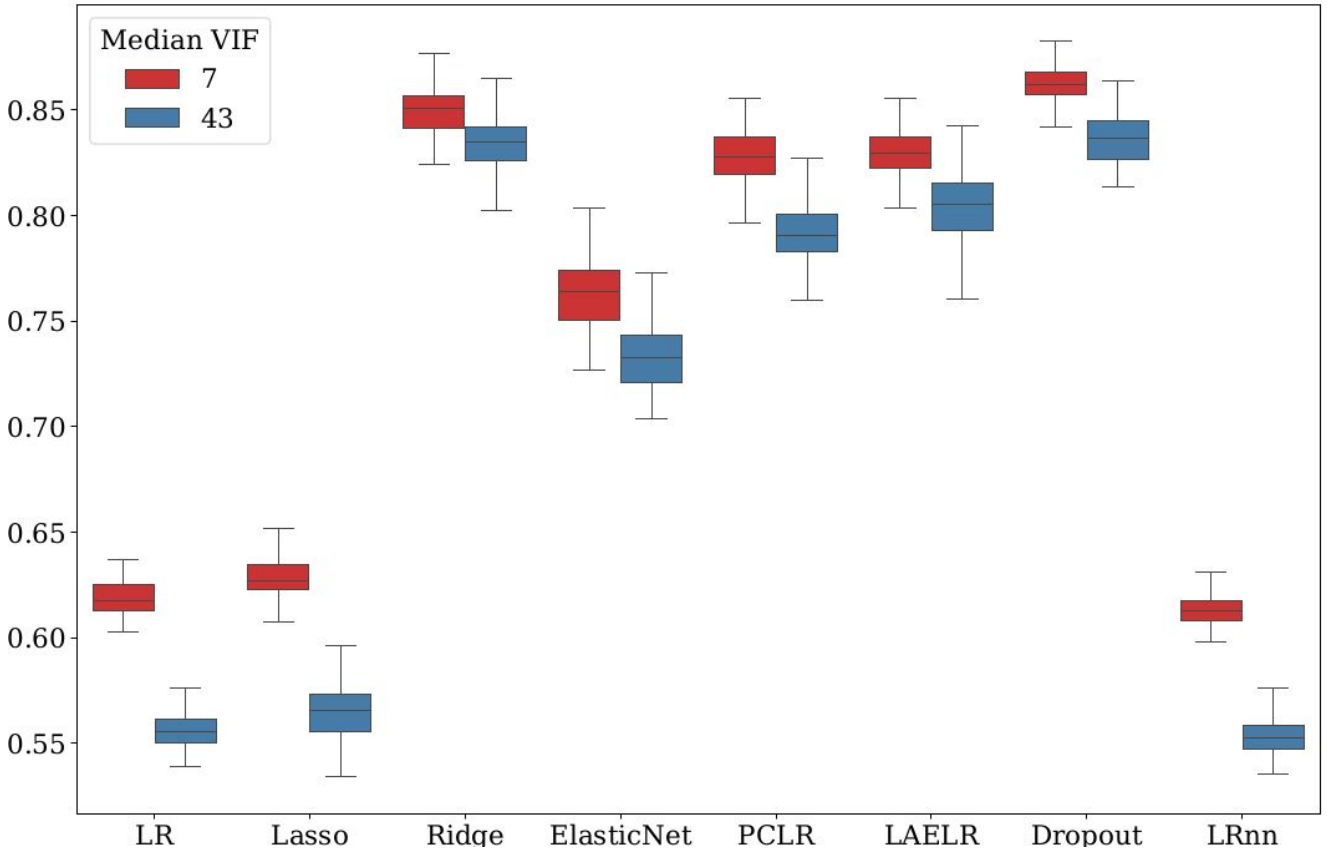


Expected Percentage of Similar Coefficients

What percentage of coefficients is expected to get the same sign when repeating the experiment?

Model A	Model B
0	0.45
0.33	0.22
-0.13	0.34

Coefficient estimation

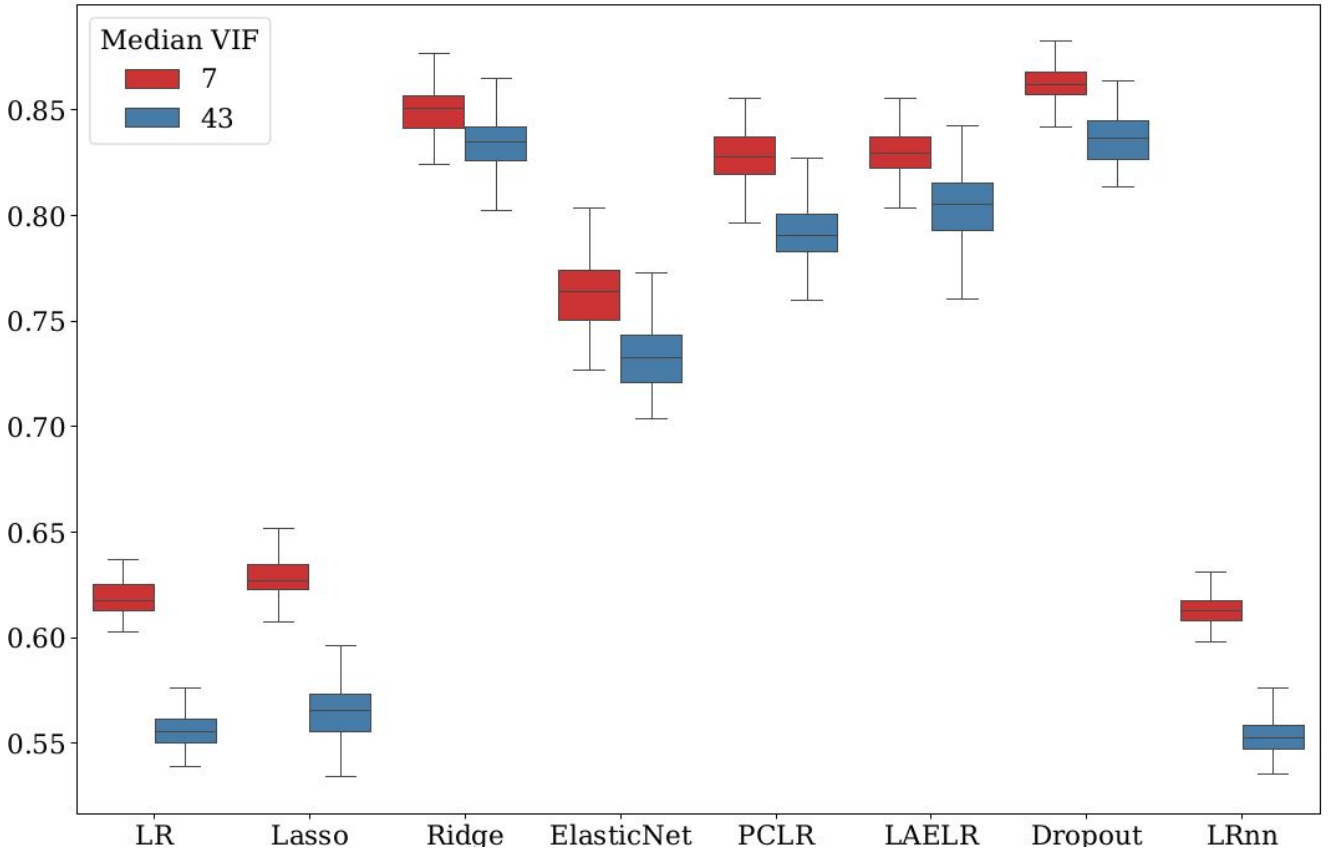


Expected Percentage of Similar Coefficients

What percentage of coefficients is expected to get the same sign when repeating the experiment?

Model A	Model B
0	+
+	+
-	+

Coefficient estimation



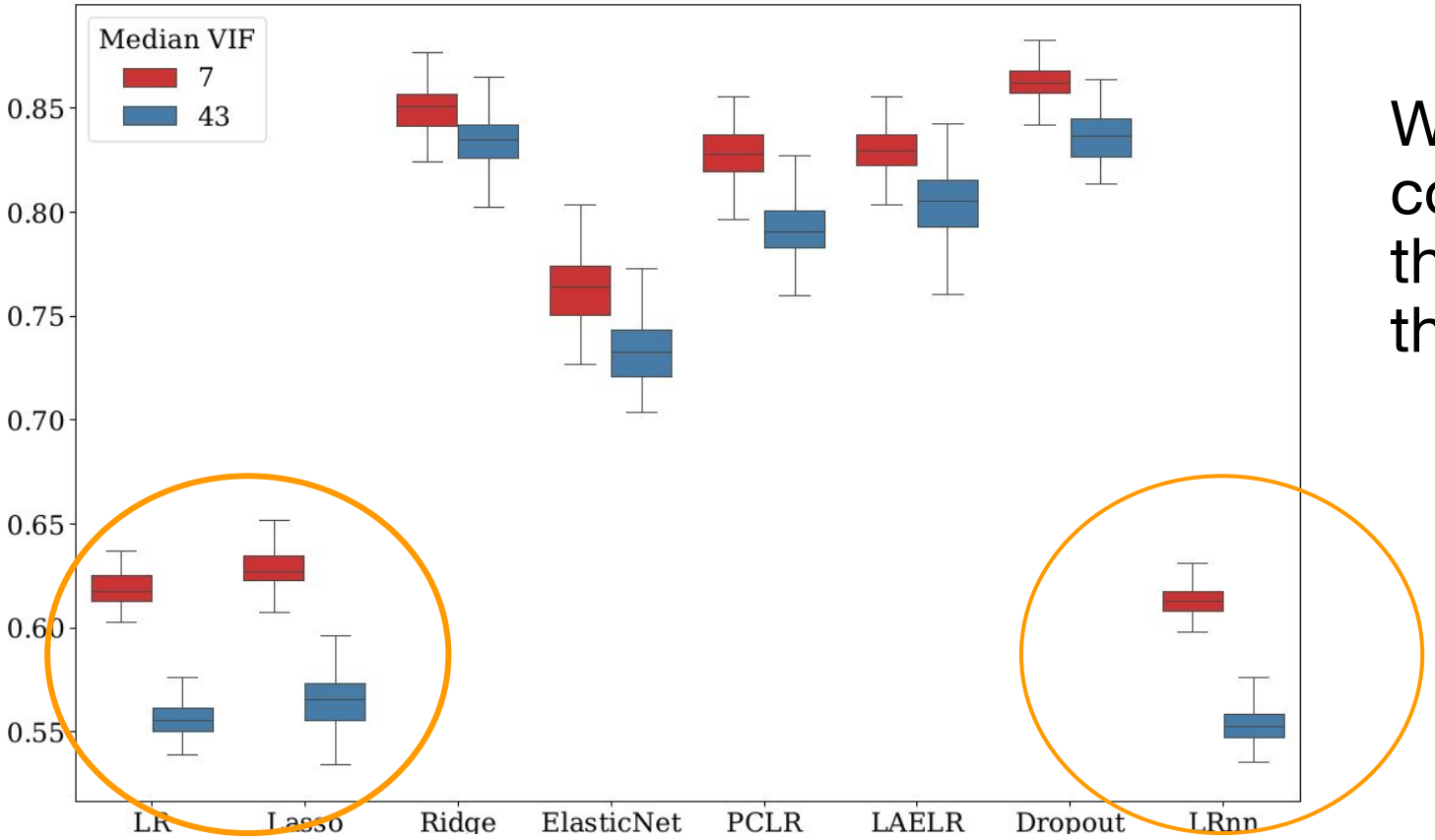
Expected Percentage of Similar Coefficients

What percentage of coefficients is expected to get the same sign when repeating the experiment?

Model A	Model B
0	+
+	+
-	+

$$\frac{1}{3}$$

Coefficient estimation



What percentage of coefficients is expected to get the same sign when repeating the experiment?

Expected Percentage of Similar Coefficients

Conclusions

(in presence of high collinearity)

<u>Predictive performance:</u>
Similar across methods.

<u>Coefficient estimation:</u>
For mostly LR, but also Lasso and LRnn

- Increase in MSE
- Less consistent selection of coefficients

Acknowledgements

Ewoud Schuit
Maarten van Smeden
Hans Reitsma
Carl Moons

Lisa van den Bosch
Arjen van der Schaaf
Hans Langendijk

 This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 825162.

Email: aleeuw15@umcutrecht.nl