Table 1: **Additional ablation studies.** "Med." is short for "Medium".

Metrics	MSE↓				MAE↓				NLL↓				AUSE↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few	All	Many	Med.	Few	All	Many	Med.	Few
VIR w/o VAE & NIG	112.62	94.21	140.03	210.72	8.18	7.44	9.52	11.45	3.787	3.689	3.912	4.234	0.451	0.460	0.399	0.565
VIR w/o NIG	87.48	73.72	107.64	161.69	7.17	6.63	8.06	9.80	3.722	3.604	3.821	4.209	0.441	0.457	0.334	0.426
VIR w/o VAE	96.46	86.72	102.56	171.52	7.51	7.08	7.93	10.45	3.784	3.685	3.866	4.218	0.439	0.459	0.385	0.423
VIR (OURS)	**81.76**	**70.61**	**91.47**	**142.36**	**6.99**	**6.39**	**7.47**	**9.51**	**3.703**	**3.598**	**3.805**	**4.196**	**0.434**	**0.456**	**0.324**	**0.414**

Table 2: **Comparison for different numbers of bins.** "Med." is short for "Medium".

Metrics	Bins	MSE↓				MAE↓				GM↓			
Shot	#	All	Many	Med.	Few	All	Many	Med.	Few	All	Many	Med.	Few
RANKSIM	100	83.51	71.99	99.14	149.05	7.02	6.49	7.84	9.68	4.53	4.13	5.37	6.89
VIR (OURS)	100	**81.76**	**70.61**	**91.47**	**142.36**	**6.99**	**6.39**	**7.47**	**9.51**	**4.41**	**4.07**	**5.05**	**6.23**
RANKSIM	33	109.45	91.78	128.10	187.13	7.46	6.94	8.42	10.66	5.13	4.70	5.23	8.21
VIR (OURS)	33	**84.77**	**77.29**	**95.66**	**125.33**	**7.01**	**6.70**	**7.45**	**8.74**	**4.36**	**4.20**	**4.73**	**4.94**
RANKSIM	20	98.71	84.38	107.89	171.04	7.32	6.78	8.35	10.57	5.33	4.51	5.69	7.92
VIR (OURS)	20	**84.05**	**72.12**	**100.49**	**151.25**	**7.06**	**6.50**	**7.90**	**10.06**	**4.49**	**4.05**	**5.34**	**7.28**

Table 3: **Comparison with DIR + Deep Ensemble.** "Med." is short for "Medium".

Metrics	MSE↓				MAE↓				NLL↓				AUSE↓			
Shot	All	Many	Med.	Few	All	Many	Med.	Few	All	Many	Med.	Few	All	Many	Med.	Few
DIR + DEEP ENS.	94.10	80.24	109.45	182.52	7.51	6.94	8.25	10.83	5.069	4.772	4.574	5.236	0.463	0.471	0.424	0.438
VIR (OURS)	**81.76**	**70.61**	**91.47**	**142.36**	**6.99**	**6.39**	**7.47**	**9.51**	**3.703**	**3.598**	**3.805**	**4.196**	**0.434**	**0.456**	**0.324**	**0.414**

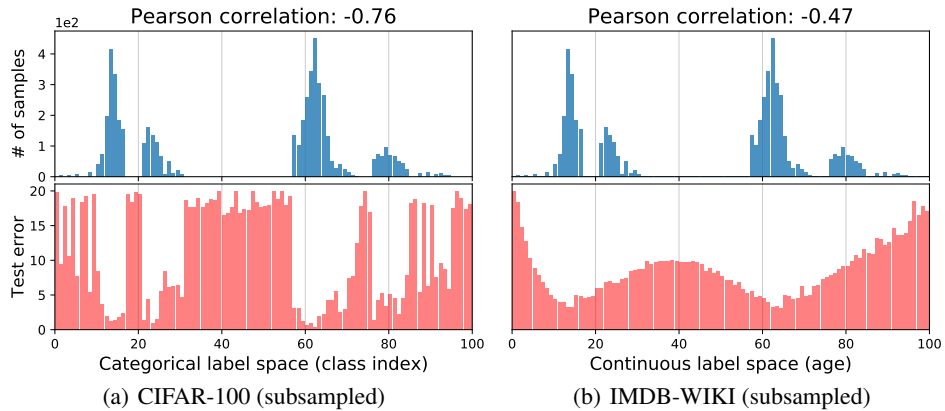(a) CIFAR-100 (subsampled)   (b) IMDB-WIKI (subsampled)

Figure 1: Comparison on the test error distribution (bottom) using same training label distribution (top) on two different datasets [1]: (a) CIFAR-100, a classification task with categorical label space. (b) IMDB-WIKI, a regression task with continuous label space.

References

[1] Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi. Delving into deep imbalanced regression. PMLR, 2021.