Supplement to Author Rebuttal

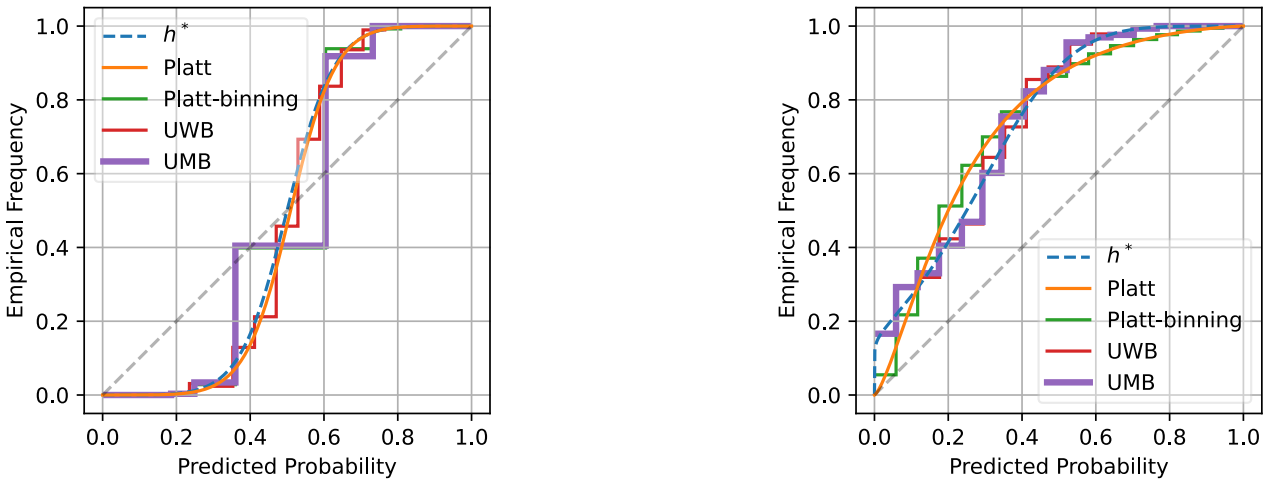
Table R.1: Quadrature estimates of population risks for Platt scaling [34], Platt-binning [26], Uniform Width Binning (UWB) [16], and Uniform Mass Binning (UMB) (**ours**) for (a) Logistic calibration and (b) Beta calibration. In (a), where the parametric assumption of Logistic model is correct, Platt-binning achieves lower R^{cal} than both the scaling method (Platt) and the binning methods (UWB and UMB). In (b), where the parametric assumption does not hold, the nonparametric binning methods outperform the parametric (Platt scaling) and hybrid (Platt-binning) methods. For both models, we sample $n = 5000$ data points (Y, Z) and choose $B = 17 \approx n^{1/3}$ number of bins for all binning schemes.

(a) Logistic calibration: $\mu_{\text{Logistic}}(z; \gamma, \delta) := \mathbb{E}[Y \mid Z = z] = 1/\left(1 + 1/e^{\gamma z + \delta}\right)$. We sample (Y, Z) from a mixture of Gaussian: $Y \sim \text{Bernoulli}(0.5)$, $Z \mid Y = 0 \sim N(-2, 1)$ and $Z \mid Y = 1 \sim N(2, 1)$.

Metric	R^{cal}	R^{sha}	R	MSE
Platt	**0.00006**	**0.00000**	**0.00006**	**0.00258**
Platt-binning	0.00006	0.00388	0.00394	0.00645
UWB	0.00015	0.00019	0.00035	0.00286
UMB	0.00007	0.00388	0.00395	0.00646

(b) Beta calibration: $\mu_{\text{Beta}}(z; a, b, c) := \mathbb{E}[Y \mid Z = z] = 1/\left(1 + 1/e^{c}\frac{z^a}{(1-z)^b}\right)$. We sample (Y, Z) following $Z \sim \text{Uniform}[0, 1]$ and $Y \mid Z \sim \text{Bernoulli}\left(\mu_{\text{Beta}}(z, a, b, c)\right)$, where $a = 1$, $b = 1$, and $c = \log 3$.

Metric	R^{cal}	R^{sha}	R	MSE
Platt	0.00299	**0.00000**	0.00299	0.02391
Platt-binning	0.00287	0.00035	0.00322	0.02413
UWB	**0.00043**	0.00036	**0.00079**	**0.02171**
UMB	0.00045	0.00035	0.00080	0.02171



(a) Logistic calibration

(b) Beta calibration

Figure R.1: Optimal recalibration function h^* and recalibration function estimates by Platt Scaling [34], Platt-binning [26], UWB [16], and UMB (**ours**) for (a) Logistic calibration and (b) Beta calibration. In (a), UMB traces the h^*, whereas scaling-binning traces Platt scaling, exhibiting an intrinsic bias from h^* in (b).