Table 1: Models used throughout the paper. D denotes the data used for the model which in our applications would be $D \in \mathcal{D} = 2^{\{\boldsymbol{x}, \mathsf{B}\}}$. Here, 2^A denotes the power set of A.

Experiment	**Name**	**Abbreviation**	**Trafo** $h(y\|D)$
Wine	Multiclass classification	MCC	
	polr	polr	$\vartheta_k - \boldsymbol{x}^\top \boldsymbol{\beta}$
	Complex intercept	CI_x	$\vartheta_k(\boldsymbol{x})$
	Simple intercept + linear shift	SI-LS_x	$\vartheta_k - \boldsymbol{x}^\top \boldsymbol{\beta}$
DR & Face	Multiclass classification	MCC	
	Multiclass classification + tabular	MCC-x	
	QWK loss	QWK	
	QWK loss + tabular	QWK-x	
	Simple intercept + tabular	SI-LS_x	$\vartheta_k - \boldsymbol{x}^\top \boldsymbol{\beta}$
	Complex intercept	CI_B	$\vartheta_k(\mathsf{B})$
	Complex intercept + tabular	$\text{CI}_\mathsf{B}\text{-LS}_x$	$\vartheta_k(\mathsf{B}) - \boldsymbol{x}^\top \boldsymbol{\beta}$
	Simple intercept + complex shift	SI-CS_B	$\vartheta_k - \eta(\mathsf{B})$
	Simple intercept + complex shift + tabular	$\text{SI-CS}_\mathsf{B}\text{-LS}_x$	$\vartheta_k - \eta(\mathsf{B}) - \boldsymbol{x}^\top \boldsymbol{\beta}$