A. Additional Results

We provide some additional results in table 2 on publicly available data sets taken from the UCI machine learning repository. Specifically we take three data sets: *Breast Cancer*, *Iris*, and *Wine* before slightly adapting them to more naturally fit the covariate shifted setting. First we make them a binary classification problem by taking the class with the largest members as positive and all others as negative. We then split the data into training and testing sets by projecting on to the first principal component and sampling a 20% testing set weighted by this value.

For all of the neural networks we consider the same architecture of two fully connected hidden layers of 32 and 64 hidden units each with tanh activation function. The initial weights are randomly drawn from N(0, 0.1) and all networks are trained using Adam. We consider the prediction accuracy as measured by AUROC shown as "TEST PERF." as well as the standard deviation of the posterior predictive distribution as a (unnormalised) predictor for whether or not the model will make an error on a given input. The corresponding AUROC score ("ERROR PRED") measures the agreement between model uncertainty and the chance to predict wrongly, and hence reflects whether the model is well-calibrated.

We see tranductive dropout always performs strongly on test performance, and though not always the best is certainly competitive in all cases, demonstrating their doesn't appear to be a toll on mean predicitive power. Further though we see that transductive dropout does remain the best across the data sets on the task of error prediction, demonstrating better uncertainty calibration, the focus of this work.

Table 2. For the three datasets we present the area under the ROC curve for two tasks, first correctly predicting the classification in the test set and secondly predicting whether for a given test point the model will make an error.

METHOD	BREAST CANCER		IRIS		WINE	
	TEST PERF.	ERROR PRED.	TEST PERF.	ERROR PRED.	TEST PERF.	ERROR PRED.
MC DROPOUT	0.979 ± 0.012	0.662 ± 0.033	0.937 ± 0.044	0.063 ± 0.046	0.972 ± 0.026	0.775 ± 0.155
CONCRETE DROPOUT	0.791 ± 0.006	0.794 ± 0.006	0.952 ± 0.038	0.847 ± 0.055	1.000 ± 0.000	0.915 ± 0.050
ENSEMBLE	0.978 ± 0.011	0.675 ± 0.007	0.960 ± 0.041	0.571 ± 0.115	0.993 ± 0.007	0.939 ± 0.037
MIXMATCH	0.986 ± 0.010	0.529 ± 0.046	0.242 ± 0.069	0.758 ± 0.069	0.889 ± 0.050	0.611 ± 0.105
LL	0.950 ± 0.033	0.329 ± 0.032	0.929 ± 0.064	0.071 ± 0.064	0.986 ± 0.013	0.575 ± 0.230
TDNR	0.979 ± 0.013	0.945 ± 0.026	0.940 ± 0.045	0.657 ± 0.105	1.000 ± 0.000	0.890 ± 0.055
TRANSDUCTIVE DROPOUT	0.968 ± 0.017	0.975 ± 0.015	0.956 ± 0.045	0.877 ± 0.082	1.000 ± 0.000	0.951 ± 0.034