Understanding **Risk Extrapolation** and when it finds **Invariant Relationships**

Delft University of Technology

1 Background

We try to contribute to the problem of **Out-of-domain generalization** by finding a modal that generalizes well over unseen domains.

For example, recognizing animals ($Y = cow$) like in figure 1. The model needs to learn the **invariant features** ($X_1 = hide$) and ignore **spurious features** ($X_2 = scenery$). A confounder ($H =$ light) randomly influences a domain. The scenery change is called **domain shift**.

Risk Extrapolation[1] (REx) finds invariant relationships by equalizing the risk across training domains by:

1. Reducing training risks
2. Increasing similarity of training risks

2 Research Question

When is Rex able to learn an invariant relationship in a synthetic dataset, and when does it fail to do so?

3 Method

1. Find out what invariant relationships are and what REx and similar methods do to find them.
2. Create synthetic datasets that expose REx to 4 common domain shifts[2] (top of figure 3).
3. Perform experiments to test the influence of the number of samples and the training domains
4. Compare to Invariant Risk Minimization[3] (IRM) and Empirical Risk Minimization[4] (ERM).

Figure 1: Example training (1) and testing (2) domain

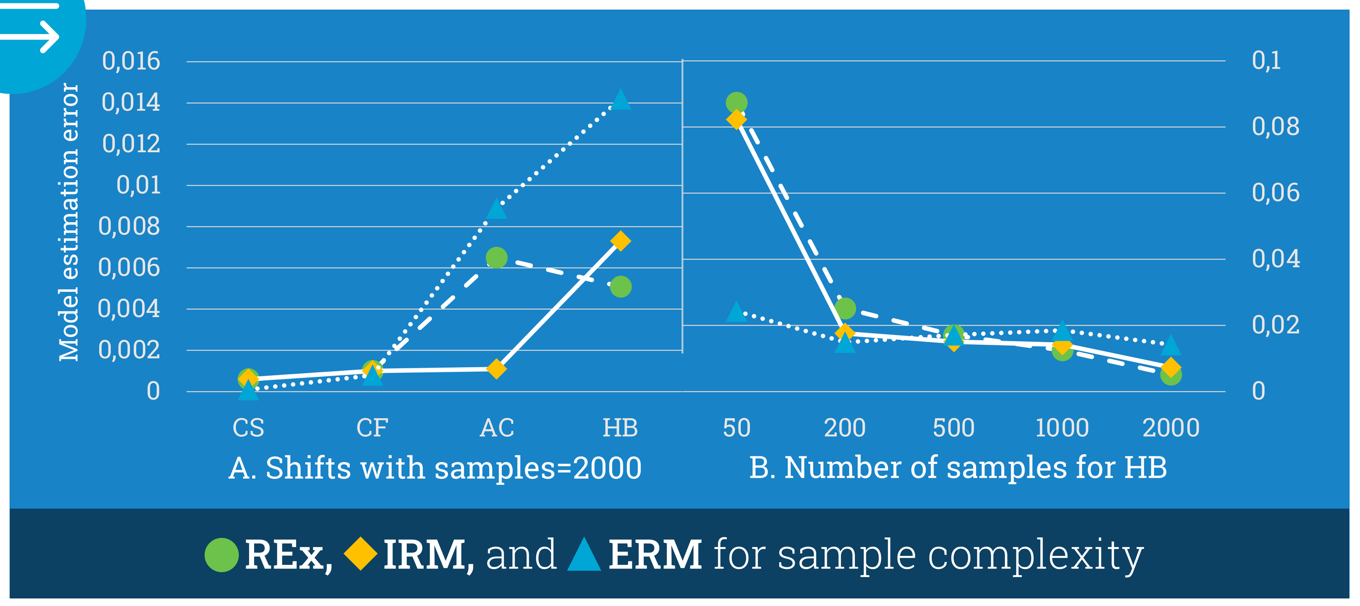
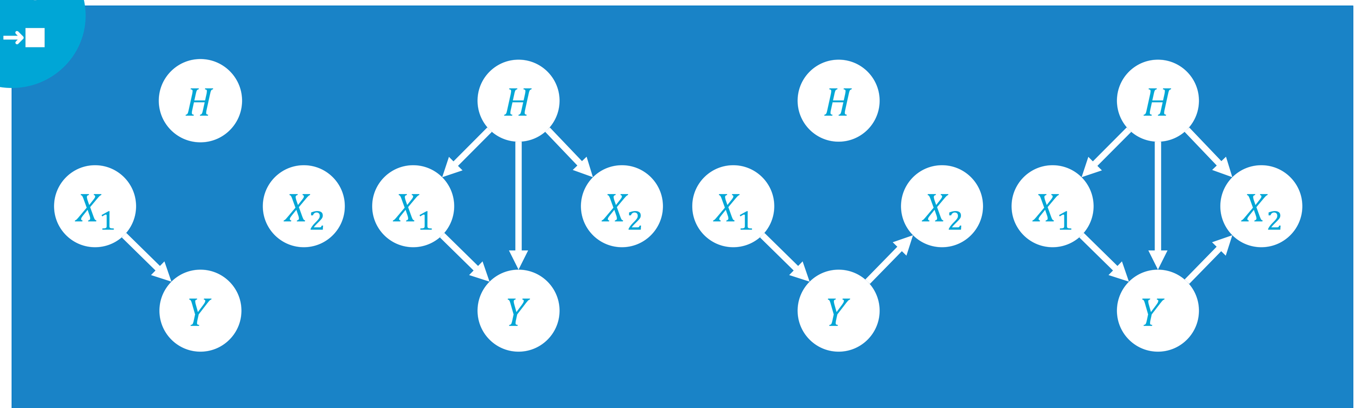


Figure 2: Sample complexity experiment



	Covariate (**CS**)	Confounded (**CF**)	Anti-causal (**AC**)	Hybrid (**HB**)
REx	0.0006±0.0001	0.0010±0.0002	0.0065±0.0013	0.0051±0.0006
IRM	0.0006±0.0001	0.0010±0.0002	0.0011±0.0002	0.0073±0.0014
ERM	0.0001±0.0000	0.0008±0.0002	0.0089±0.0002	0.0142±0.0012

Figure 3: model estimation error ± standard error for data models with 2000 samples

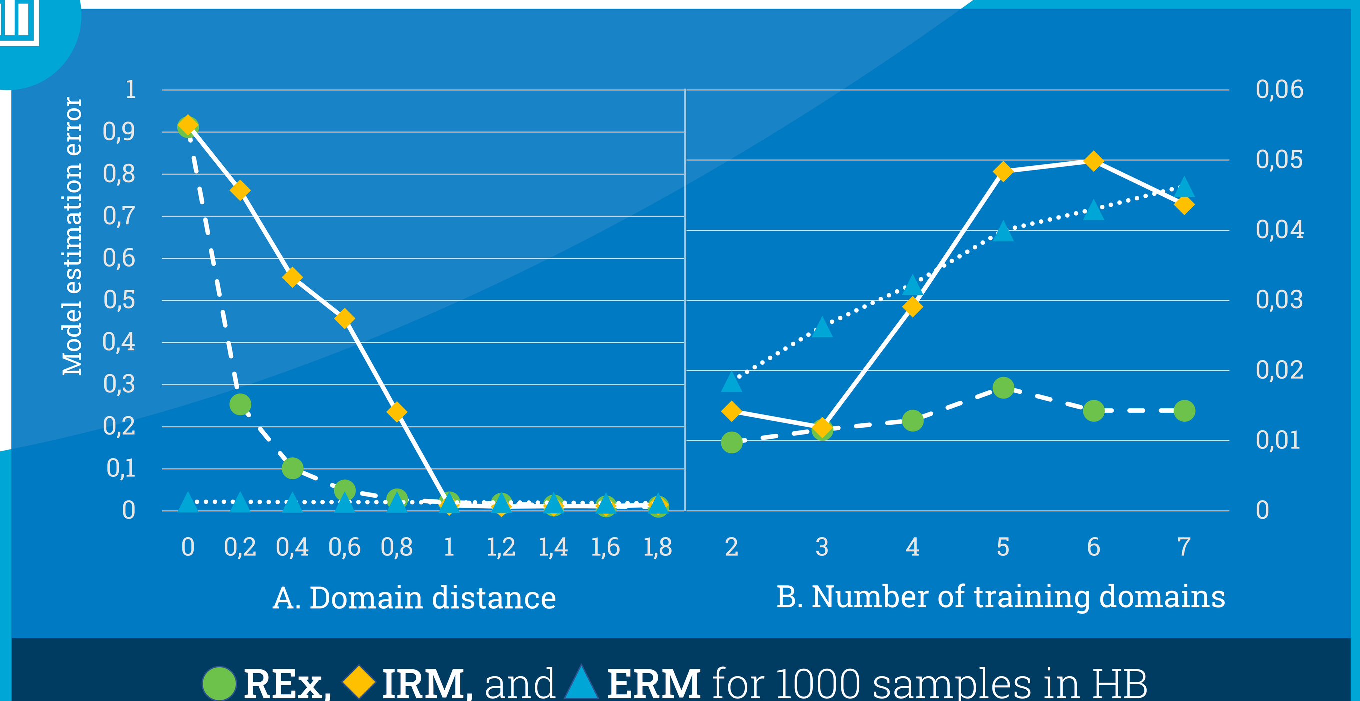


Figure 4: Domain distance and quantity of training domains experiment

4 Results

We test the methods in all 4 shifts on the sample complexity (figures 2 and 3), domain distance in figure 4A (i.e. the difference in distribution), and quantity of training domains within a fixed domain interval (figure 4B).

5 Conclusions

REx performs better for invariant prediction in situations with larger sample sizes (figure 2B) and training domain distance (figure 4A). In which the latter results in more significant differences. If these criteria are met, REx performs equivalently in all four distributional shifts (figure 3, row 1).

Compared to IRM and ERM, REx is more robust to shifting the average distributional variance in the training domains (figure 4B). Rex also asymptotically outperforms the methods in the more complex distributional shifts, such as the hybrid shift (figure 2A). The other shifts generally favour IRM and ERM. Experiments are limited by input vector size and the number of repetitions.

Jeroen Hofland – J.L.Hofland@student.tudelft.nl
Rickard Karlsson, Stephan Bongers, and Jesse Krijthe
As supervisors and responsible professor

1. Krueger, D., Caballero, E., Jacobsen, J.-H., Zhang, A., Binas, J., Zhang, D., . . . Courville, A. (2020, March). Out-of-Distribution Generalization via Risk Extrapolation (REx). *arXiv e-prints*, arXiv:2003.00688.
2. Ahuja, K., Wang, J., Dhurandhar, A., Shanmugam, K., & Varshney, K. R. (2020, October). Empirical or Invariant Risk Minimization? A Sample Complexity Perspective. *arXiv e-prints*, arXiv:2010.16412.
3. Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019, July). Invariant Risk Minimization. *arXiv e-prints*, arXiv:1907.02893.
4. Vapnik, V. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5), 988-999. doi: 10.1109/72.788640