# Test for non-negligible adverse shifts: Supplementary material

**Vathy M. Kamulete**

Enterprise Model Risk Management
Royal Bank of Canada
Toronto, Canada
vathy.kamulete@rbccm.com

## 1 EXPERIMENT DETAILS

For the empirical experiments, we use isolation forest from the `isotree` package [Cortes, 2020]. We take the default hyperparamters from `isotree` as given, but increase the ensemble size to 500. To investigate other notions of outlyingness, we use random forests from the `ranger` package [Wright and Ziegler, 2017]. We take its default hyperparameters from Probst et al. [2019].

All experiments were run on a commodity desktop computer with a 12-core Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz processor and 64 GB RAM in R version 3.6.1 (2019-07-05). We stress that no hyperparameter tuning was performed - we set the hyperparameters to reasonable defaults as previously discussed. To avoid ambiguity, we explicitly say when the results pertain to `DSOS-SS`, `DSOS-PT` or `DSOS-CV`.

### 1.1 SIMULATED SHIFTS

We simulate distribution shifts from a two-component multivariate Gaussian mixture model (GMM). The training and test set are drawn from:

$$X^{tr} \sim \phi^{tr} \cdot \mathcal{N}_d(\mu_{\mathbf{1}}^{\mathbf{tr}}, \mathbf{\Sigma}_{\mathbf{1}}^{\mathbf{tr}}) + (1-\phi^{tr}) \cdot \mathcal{N}_d(\mu_{\mathbf{2}}^{\mathbf{tr}}, \mathbf{\Sigma}_{\mathbf{2}}^{\mathbf{tr}})$$
$$X^{te} \sim \phi^{te} \cdot \mathcal{N}_d(\mu_{\mathbf{1}}^{\mathbf{te}}, \mathbf{\Sigma}_{\mathbf{1}}^{\mathbf{te}}) + (1-\phi^{te}) \cdot \mathcal{N}_d(\mu_{\mathbf{2}}^{\mathbf{te}}, \mathbf{\Sigma}_{\mathbf{2}}^{\mathbf{te}})$$

Omitting subscripts and superscripts for brevity, $\phi \in [0,1]$, $\mu$ and $\mathbf{\Sigma}$ are the component weight, mean vector and covariance matrix respectively. The baseline specifies the training and test sample size $n^{tr} = n^{te} \in \{400, 800, 1600\}$, the number of dimensions $d \in \{4, 8, 16\}$, the component weights $\phi^{tr} = \phi^{te} = 0.5$, the mean vectors $\mu_{\mathbf{1}}^{\mathbf{tr}} = \mu_{\mathbf{1}}^{\mathbf{te}} = \mathbf{1}_d$ and $\mu_{\mathbf{2}}^{\mathbf{tr}} = \mu_{\mathbf{2}}^{\mathbf{te}} = -\mathbf{1}_d$ and the covariance matrices $\mathbf{\Sigma}_{\mathbf{1}}^{\mathbf{tr}} = \mathbf{\Sigma}_{\mathbf{1}}^{\mathbf{te}} = \mathbf{\Sigma}_{\mathbf{2}}^{\mathbf{tr}} = \mathbf{\Sigma}_{\mathbf{2}}^{\mathbf{te}} = \mathbf{I}_d$. $\mathbf{I}_d$ is the $d \times d$ identity matrix and $\mathbf{1}_d$ is the $d$-dimensional all-ones vector. The baseline configurations enforce that training and test set are drawn from the same distribution, i.e. no shift. There is a total of 9 such configurations (3 dimensions times 3 sample sizes).

We generally shift the distribution so that the dimension of change $d^* = 1$ is fixed as the ambient dimension $d$ increases. The power of multivariate tests based on kernels and distances decays with increasing dimension when differences only exist along a few intrinsic dimensions $d^* \ll d$. We vary one or more parameters, namely $\phi^{tr}$, $\phi^{te}$, $\mu_{\mathbf{2}}^{\mathbf{te}}$ and $\mathbf{\Sigma}_{\mathbf{2}}^{\mathbf{te}}$, to simulate the desired shifts, all else constant. We change the following parameters to pre-set intensity levels:

1. Label (prior) shift – We flip the weights so that $\phi^{tr} \in \{0.49, 0.47, 0.45\}$ goes with $\phi^{te} = 1 - \phi^{tr}$. The majority component in training becomes the minority in the test sample.
2. Corrupted sample – We draw a fraction $\omega \in \{0.01, 0.02, 0.04\}$ of examples in the test set from the component that is absent in training such that $\phi^{tr} = 1$ and $\phi^{te} = 1 - \omega$.
3. Mean shift – We change the mean vector in the test set so that $\mu_{\mathbf{2}}^{\mathbf{te}} = [-\frac{\kappa}{10}, -\mathbf{1}_{d-1}]$, where $\kappa \in \{11, 12, 14\}$.

*Accepted for the 38th Conference on Uncertainty in Artificial Intelligence* (UAI 2022).

4. Noise shift – We change the covariance matrix in the test set so that $diag(\mathbf{\Sigma_1^{te}}) = [\frac{\theta}{10}, \mathbf{1}_{d-1}]$, where $\theta \in \{11, 12, 14\}$ and $diag(\cdot)$ is the assignment operator for the diagonal elements of the $d$-by-$d$ covariance matrix.
5. Dependency shift – We induce a positive relationship between the first two covariates. We change the covariance structure in the test set so that $\mathbf{\Sigma_2^{te}}[1, 2] = \mathbf{\Sigma_2^{te}}[2, 1] = \gamma$, where $\gamma \in \{0.1, 0.2, 0.4\}$.

There is a total of 27 configurations for each shift type. For each shift type and shift intensity, we repeat the experiment 500 times as noted in the main text.

### 1.2 PARTITION-INDUCED SHIFTS

For each OpenML-CC18 dataset, we perform stratified 10-fold cross-validation, repeated twice. We end up with 20 train-test splits per task. In total, we run 3720 tests of no adverse shifts (62 datasets, 20 random splits, and 3 tests). Summary statistics for these datasets and granular test results are bundled together in the submission package for this paper. We expect the statistical tests to be correlated within but not across datasets. To formalize this setup, we use the following model:

$$\log s_i^j \sim \mathcal{N}(\mu_i^j, \Sigma_i^j) \tag{1}$$

where for each dataset $i = 1, 2, \cdots, 62$ and test $j = 1, 2, 3$, the $s-$value $s_i^j$ is lognormally distributed: positive and skewed to the right. The $s-$value $s_i^j$ consists of a dataset-specific (fixed) effect $\mu_i^j$, subject to noise in $\Sigma_i^j$; $\Sigma_i^j$ accounts for within-dataset covariance.

We fit the model in 1 using the `clubSandwich` package to obtain robust estimates of the fixed effects $\mu_i^j$ even with arbitrary covariance structure in $\Sigma_i^j$ left unspecified [Pustejovsky and Tipton, 2018]. These fixed effects (means) $\mu_i^j$ in 1 measure how sensitive on average a dataset is to these partition-induced shifts. The higher the value, the more susceptible a dataset is to adverse shifts caused by sampling variation. As mentioned in the paper, the exponentiated means $\mu_i^j$ are on the $s$-value scale and can be interpreted as the strength of evidence against the null of no adverse shift.

### References


David Cortes. *isotree: Isolation-Based Outlier Detection*, 2020. URL `https://CRAN.R-project.org/package=isotree`. R package version 0.1.18.

Philipp Probst, Anne-Laure Boulesteix, and Bernd Bischl. Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*, 20(53):1–32, 2019.

James E Pustejovsky and Elizabeth Tipton. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4):672–683, 2018.

Marvin N. Wright and Andreas Ziegler. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1):1–17, 2017. doi: 10.18637/jss.v077.i01.