Code and data for reproducing the empirical results in the paper "Regression with linked data sets subject to linkage error"

This directory is divided into five main subdirectories.

1. Subdirectory `table2` contains MATLAB code and result files (`.mat`) that reproduce the results in table 2 of the paper.

2. Subdirectory `figure3` contains MATLAB code and result files (`.mat`) reproducing the simulation results in figure 3. The robust approach in Slawski and Ben-David (2019) requires robustfit function in matlab.

3. Subdirectory `figure4` contains MATLAB code and result files (`.mat`) reproducing the simulation results in figure 4.

4. Subdirectory `data` contains data files reproducing the results in table 2, figure 3 and figure 4.

5. Subdirectory `functions` contains MATLAB code for specific algorithms and approaches in the paper.

Guide for the subdirectory `table2`, `figure3` and `figure4`

1. Subdirectory contains three folders with `.mat` files that can contains result of running the code with prefix `run`. For `figure3`, the result is stored as `x-y.mat`.

2. Table 2 result are stored as `error-est` and `r-square` which correspond to relative error and R^2 respectively.

3. For `table2`, running END takes long time due to large sample size.

Guide for the subdirectory `data`

1. The path for MATLAB code is already been set up in each code with prefix `run`, no need to add data into each folder before running the code

Guide for the subdirectory `functions`

1. Prefix `EM` stands for EM algorithm

2. Prefix `DA` stands for Data Augmentation

3. Suffix `mixture` stands for the mixture approach in Slawski, Diao and Ben-David (2020)

4. There are two mex file in the folder which is created by mex MATLAB 2019a.

To regenerate the results in the three subdirectories, call the files with prefix run, which in turn automatically call the corresponding files with prefix experiment. Note that execution of the latter will over-ride all existing .mat files.