Results for simulation

3_predictors non paper

- the test tries to estimate the coefficients of "simulations\settings\3_predictors\default.R"
- notice that the simulation type is "cov", so the X data is generated adding a covariate error over each observations for the same feature.
- simple_regressor MIP -> the OLS solution is very close to the MIP solution but is not the same
- regressor_with_group_constraint MIP -> the OLS solution is very close to the MIP solution but is not the same
- the MIP solution seems better than the OLS one, the performance metrics are better
    - coefficients are closer to expected ones
    - prediction errors on the training set (Y, Z) are lower than using OLS solutions
    - prediction errors are also lower than using the EXPECTED BETA coefficients -> the MIP solver overfitted the data
- Just run the 3_predictors_non_paper_simulation and change the MODEL name
-  Alt text

5_predictors non paper

- the test tries to estimate the coefficients of "simulations\settings\5_predictors\default.R"
- notice that the simulation type is "cov", so the X data is generated adding a covariate error over each observations for the same feature.
- simple_regressor MIP -> the OLS solution is very close to the MIP solution but is not the same
- regressor_with_group_constraint MIP -> the OLS solution is very close to the MIP solution but is not the same
- the MIP solution seems better than the OLS one, the performance metrics are better
    - coefficients are closer to expected ones
    - prediction errors on the training set (Y, Z) are lower than using OLS solutions
    - prediction errors are also lower than using the EXPECTED BETA coefficients -> the MIP solver overfitted the data
- Just run the 3_predictors_non_paper_simulation and change the MODEL name
-  Alt text
- Params: 150 observations were not enough. With 350 the results improved a lot!

Zambon PAPER

- the test tries to estimate the coefficients of "simulations\settings\paper\default.R"
- if BIG M bounds are not set properly, for example M= +/- 300000, the model is not able to do feature selection and selects the wrong predictors with bad values.
- The OLS solution for this problem has also very big coefficients
   Alt text
- While if we reduce the value of Big M we can obtain a more correct solution (feature selection wise) but incorrect coefficients. i.e. M= +/- 10
-  Alt text
- Rule of thumb: if gurobi sets a beta coefficients to the value of Big-M, lower the Big-M bound value...Repeat! (that's what internet folks say..)
- The difficulty here is that each predictor is defined over its own "TIME DOMAIN". Beta coeff could vary highly because of this reason.
- Notice that the "pre processing steps" required this assumption "the betas are defined over the same time domain as their associated predictor curve".
- Even by lowering BIG_Ms to 1, we are not able to find the exact curves
- Another test that was done was "reducing the basis functions degree" to only 4 (instead of 6). That combned with the "big ms set to +1/-1" improved the results and 2 of the 3 predictors curves were fitted.

Gertheissa PAPER

- the test tries to estimate the coefficients of "simulations\settings\paper2\default.R"
- if BIG M bounds are not set properly, for example M= +/- 300000, the model is not able to do feature selection and selects the wrong predictors with bad values.
- Even if BIG M is properly set (+/- 1), we still need to provide a lot of observations to better fit the curves. For example:
   Alt text