

I am concerned about how much overfitting is being introduced into the data by adding and selecting responses and adding predictors, many of which use information from the entire period up to era 574. Initially, if information from eras after 574 are not used, everything seems correct, but the testing and selection process of promising predictors and targets itself introduces a selection bias that generates overfitting. I hope Numerai has taken all of this into account.