

## Code and Data Guide

I would like to thank Peter Winker and the Journal of Economics and Statistics for accepting my paper "Testing Investment Forecast Efficiency with Forecasting Narratives." Below you find a brief description of the data and code I used to create the paper. If the folder structure is contained, all programs should operate after installing the required packages and producing my paper's results. It is required to use the code in RStudio with the package rstudioapi installed. For questions, please contact me at [foltasa@hsu-hh.de](mailto:foltasa@hsu-hh.de).

### Data

The folder "data\_tm" contains the textual accounts used to create the topic models. The folder "data\_rf" includes the numerical point forecasts of the investigated institutes (data.forecast.xlsx) and the used indicators (all\_indicators.xls; tab "indicators\_with\_publ\_lag").

### Code

The folder "programs\_tm" contains all programs related to used topic models, including the developed topic model in "Word2Sense-LDA" and a standard LDA model in folder "LDA." The folder "Topic quality" contains the program that was used to create Figure 2.

The folder "programs\_fe" contains all programs related to the forecast efficiency analysis. The folder "programs\_rf" contains all programs related to the forecast efficiency analysis. The folder "data.compute" reshaped the corpus while "random forests" contains all random forest analyses sorted into subfolders for different settings. For replicating my results, it is required to start the program "data.rf.topics" and subsequently the "collect\_results" program. The "visualization" folder contains additional programs concerning the tables ("export\_summary" and "export\_results"), the %Inc graphs ("export\_%incMSE\_ap") and the pdp graphs ("export\_pdp\_mean"). For working with these programs, it is necessary to define the dir parameter to a subfolder of the "random\_forests" folder.