**Seminar Case Studies - To Do List 01.02.23**

Programming

* FM Model: path-dependent vs direct? check with team whether all fine
* SSA implementation + running predictions
* Running random/boosted forests with correct MAF (?)
* Variable importance: graphs?
* Graphs: what do we need? Settle on graph layout
* Tables: what else do we need to do to be able to fill in our tables?
* Once Coulombe confirmed, run predictions with Dutch data
* Once Coulombe confirmed, run predictions with Dutch data + additional stuff

Writing

* Literature
  + Random/boosted forests, other concepts/methods used throughout
  + Aim here is to add 1/2 pages
  + Trace back references in Coulombe paper
* Data:
  + Argue why we add certain data, perhaps papers arguing why (stats journal)
  + Augmented Phillips curve
  + Anything else we should add here? Rather short now...
  + Perhaps explain how we calculate growth?
* Methodology
  + 4.3. becomes "Forecasting Methods", rather than ML Methods?
  + In 4.3 explain how Random Forests/Boosted Trees/FM "tuned"
  + Explain how we measure performance; relative: FM model is benchmark
  + If/when: macroeconomic random forests (then also include in results)
* Results
  + Give overview of results for FM, Random, Boosted for Direct and Path-Dep
  + Repeat analysis with/without SSA - improvement?
  + Report optimal configuration for every horizon/direct-pathav.
  + Report variable importance and R^2-regression results (like Coulombe)
* Conclusion
  + Compare results to Coulombe; do data transformations matter for Dutch data?
  + If so, which transformations are most important, for all horizons?
  + Next steps; what would be interesting to research to complement our project?

Tables

* Relative RMSE; horizons (columns), feature matrices (rows); For direct and path average, for trees and boosted

Figures

* Marginal contribution features; path-average/direct, forests/boosted
* Variable importance; path-average/direct, random forest