Data Guide for "Forward or Backward Looking? The Economic Discourse and the Observed Reality"

forthcoming in Journal of Economics and Statistics 236,4

Jochen Lüdering Peter Winker

June 24, 2016

This document gives a brief description of the data supplied along with our manuscript. Using the R and EViews the results of our paper can be replicated. Due to the size of the corpus we are not supplying the orgininal corpus but only the derived Document–Termmatrix. If you are interested in working with the corpus please get in touch.

The corpus consists of all volumes published between 1949 and 2010. From the original documents we removed footnotes and the table of contents of the individual issues. Further, we striped the articles of all tables, equations, the bibliography and the appendix unless it included only text. Afterwards we proceeded with the preparation as stated in our manuscript. We applied a stemming algorithm and removed all superfluous blanks, newline and tabulator codes, as well as punctuation and digits. Afterwards we removed all German stop words and only considered terms between 5 and 20 characters in length. In the next step we created the Document-Term Matrix and reduced the vocabulary by apply selecting the terms based on their *tf-idf* scores. These steps result in the document-termmatrix provided here.

Contents

The data packages consists of five files.

jbnst-lda.R This R script takes a DocumentTermmatrix and fits a topic model. It will save the topic probalities as an *.xlsx* file (for illustrative purposes). It also saves the aggregated topic probabilities into *dataset.xlsx*.

FinalTermmatrix.robj The Document–Termmatrix used by *jbnst-lda.R*.

metadata.robj Meta data used by *jbnst-lda.R*. The meta data was obtained from `digizeitschriften.de` and manually corrected by checking the titlepages were information appeared implausible.

regression.prg An EViews program running the univariate and multivariate regressions in the paper.

regressionvars.xlsx Input data for the Eviews program. It includes the real world time-series and the topic probabilities from our paper.

How to

- Run the topic model (*jbnst-lda.R*) in R with the necessary packages installed.
- Gibbs Sampling involves randomness. While the composition of the topics is robust, the order of the generated topics might be different. Thus, manually identify the correct topics and copy and use the corresponding probabilities in the model (e.g. copy & paste into *regressionvars.xlsx*). **This is manual work!**.
- Run *regression.prg* in Eviews. Unless you changed the probabilities you are using our original topic probabilities.