

Zeit Index

Dirk Ulbricht

Thursday, January 08, 2015

Description

The German weekly journal “Die Zeit” from January 1990 to Dezember 2014 has been downloaded from its publicly accessible online archive. As the indeces to be constructed are intended to be informative of the business cycle, only the economics, politics and opion sections that are supposed to be relevant to this end are analysed. Additional to the construction of indeces based on all three sections, a corresponding set of indeces is constructed based on the economics section only (the corresponding indicators are labeled “economics”).

The calculation of the indeces is based on the results of a preceding sentiment analysis. Each word is matched to the terms comprised in SentiWS, a dictionary of positive and negative words, that attributes numerical values to each term indicating the degree of sentiment they convey.

	value	words
2	0.004	Abschluß,Abschlüße,Abschlußs,Abschlußes,Abschlüssen,
3	0.004	Abstimmung,Abstimmungen,
4	0.004	Agilität,
5	0.004	Aktivität,Aktivitäten,
6	0.004	Aktualisierung,Aktualisierungen,
7	0.004	Aktualität,Aktualitäten,

Table 1: Example SentiWS

In a first step the number of positive and negative words and the sum of positive and negative sentiment values as well as the total number of words are collected on an article basis. In turn, two groups of indicators are computed.

Number of words indicators: The first set of indicators is only based on the counts of positive and negative words and ignores sentiment values. As the article size varies, they are calculated as ratios of the total number of words of each article. The ratio of the difference and the sum of the number of positive and negative words (perc_dif and perc_sum) to the total number of words and the number of positive and negative words to the total number of words (perc_pword and perc_nword) are calculated.

Sentiment value indicators: The second set of indicators is based on the sentiment values of each article. The variables rpvalue and rnvalue label indicators based on averages of positive and negative values, respectively. The ratio of the balance of positive and negative values as a ratio of the sum of positive and negative words is labeled rvalue.

For each of the two groups, the monthly averages are calculated. As a measure of uncertainty, the variances (labeled “var”) and as a measure of risk the kurtosis (labeled “kurt”), both on a monthly basis, are computed for both sets of indicators.

For all of the above indicators, a three month moving average (labeled “3”) is calculated.

label	description
perc_dif	average balance of the number of positive and negative words
perc_sum	average of the number of positive and negative words
perc_pword	number of positive words divided by the total number of words
perc_nword	number of negative words divided by the total number of words
pvalue	average positive value
nvalue	average negative value
rvalue	average balance of positive and negative values
var	variance based on values of all articles each month
kurt	kurtosis based on values of all articles each month
3	three month moving average
economic	values based on articles in the economics section only

Table 2: label description