Vignette ifwtrends

30.8.2021

Funktionen

pca

Die Funktion `pca` nimmt als Argumente mehrere Suchwörter oder Kategorien entgegen. Des weiteren eine Region, das Start- und Enddatum (Default: 2006-01-01 und heute) sowie die Anzahl der zu berechnenden Hauptkomponenten (Default: Anzahl der Zeireihen). Für die Zeitreihen wird hier momentan monatliche Frequenz angenommen.

```
pca(keywords = c("ikea", "saturn"),
    categories = 0,
    geo = "DE",
    start = "2006-01-01",
    end = Sys.Date(),
    components = max(length(keywords), length(categories)))
#> # A tibble: 188 x 5
#>    date         PC1    PC2  ikea saturn
#>    <date>     <dbl>  <dbl> <int>  <int>
#>  1 2006-01-01 -27.4  9.61     25     49
#>  2 2006-02-01 -37.6 -4.20     23     32
#>  3 2006-03-01 -36.1 -1.59     23     35
#>  4 2006-04-01 -34.5  3.26     22     40
#>  5 2006-05-01 -37.4 -1.97     22     34
#>  6 2006-06-01 -35.0  2.39     22     39
#>  7 2006-07-01 -35.1  0.152    23     37
#>  8 2006-08-01 -27.5 -0.684    30     40
#>  9 2006-09-01 -29.8 -0.570    28     39
#> 10 2006-10-01 -23.6  6.28     30     48
#> # ... with 178 more rows

pca(keywords = NA,
    categories = c(651),
    geo = "DE",
    start = "2006-01-01",
    end = Sys.Date(),
    components = max(length(keywords), length(categories)))
#> # A tibble: 188 x 3
#>    date         PC1 `651`
#>    <date>     <dbl> <int>
#>  1 2006-01-01  40.0    85
#>  2 2006-02-01  25.0    70
#>  3 2006-03-01  30.0    75
#>  4 2006-04-01  23.0    68
#>  5 2006-05-01  34.0    79
#>  6 2006-06-01  27.0    72
```

```
#>  7 2006-07-01  31.0    76
#>  8 2006-08-01  29.0    74
#>  9 2006-09-01  24.0    69
#> 10 2006-10-01  20.0    65
#> # ... with 178 more rows
```

roll

Roll gibt pca für jeden Zeitpunkt in einem Zeitfenster neu an. Dies gibt eine Tabelle mit den Zeitpunkten im Zeitfenster `start_period` bis `end` zurück. In jeder Spalte ist dann die Ausgabe von pca, wobei `end` der jeweilige Zeitpunkt im Zeitfenster ist. Die restlichen Zeilen zum Tabellenende sind `NA`s. Dies kann insb zur Prognoseevaluation genutzt werden. Sowohl `pca` als auch `roll` könnten bei Bedarf umgeschrieben werden, sodass statt den Hauptkomponenten andere Faktoren berechnet werden

```
roll(keywords = "ikea",
     geo = "DE",
     start_series = "2006-01-01",
     start_period = "2006-05-01",
     end = "2006-12-01")
#> # A tibble: 12 x 17
#>    date       `PC1 to 2006-05-01` `ikea to 2006-05-01` `PC1 to 2006-06-01`
#>    <date>                   <dbl>                <int>                <dbl>
#>  1 2006-01-01                   2                   25                 2.17
#>  2 2006-02-01                   0                   23                 0.167
#>  3 2006-03-01                   0                   23                 0.167
#>  4 2006-04-01                  -1                   22                -0.833
#>  5 2006-05-01                  -1                   22                -0.833
#>  6 2006-06-01                  NA                   NA                -0.833
#>  7 2006-07-01                  NA                   NA                NA
#>  8 2006-08-01                  NA                   NA                NA
#>  9 2006-09-01                  NA                   NA                NA
#> 10 2006-10-01                  NA                   NA                NA
#> 11 2006-11-01                  NA                   NA                NA
#> 12 2006-12-01                  NA                   NA                NA
#> # ... with 13 more variables: ikea to 2006-06-01 <int>,
#> #   PC1 to 2006-07-01 <dbl>, ikea to 2006-07-01 <int>, PC1 to 2006-08-01 <dbl>,
#> #   ikea to 2006-08-01 <int>, PC1 to 2006-09-01 <dbl>,
#> #   ikea to 2006-09-01 <int>, PC1 to 2006-10-01 <dbl>,
#> #   ikea to 2006-10-01 <int>, PC1 to 2006-11-01 <dbl>,
#> #   ikea to 2006-11-01 <int>, PC1 to 2006-12-01 <dbl>, ikea to 2006-12-01 <int>
```

daily_series

Für lange Zeitfenster liegen keine täglichen Daten vor sondern nur monatliche. Die Funktion `daily_series` zieht zunächst für rollierende Zeiträume mehrere Stichproben und schätzt daraus dann mit Chow-Lin für den ganzen Zeitraum tägliche Daten. Diese sind konsistent mit den Monatsdaten. Da momentan sehr viele Samples gezogen werden, verursacht die Funktion viele Suchanfragen bei Google, was nach einigen malen zur vorübergehenden Sperrung der IP führt. Die Anzahl der gezogenen Fenster ist momentan unter der aus dem Originalcode um Anfragen zu sparen. Die dadurch hervorgerufene Abweichung scheint im Moment sehr gering bis 0 zu sein. Eine genaue Evaluierung steht hier aber noch aus, ist jedoch wegen dem noch nicht gelösten IP-Problem momentan nicht durchführbar.

```
daily_series(keyword = c("arbeitslos"),
             geo = "DE",
```

```
             from = "2021-01-01")
#> # A tibble: 239 x 2
#>    time       value
#>    <date>     <dbl>
#>  1 2021-01-01  33.0
#>  2 2021-01-02  30.0
#>  3 2021-01-03  71.0
#>  4 2021-01-04  55.0
#>  5 2021-01-05  54.0
#>  6 2021-01-06  61.0
#>  7 2021-01-07  52.0
#>  8 2021-01-08  49.0
#>  9 2021-01-09  38.0
#> 10 2021-01-10  13.0
#> # ... with 229 more rows
```

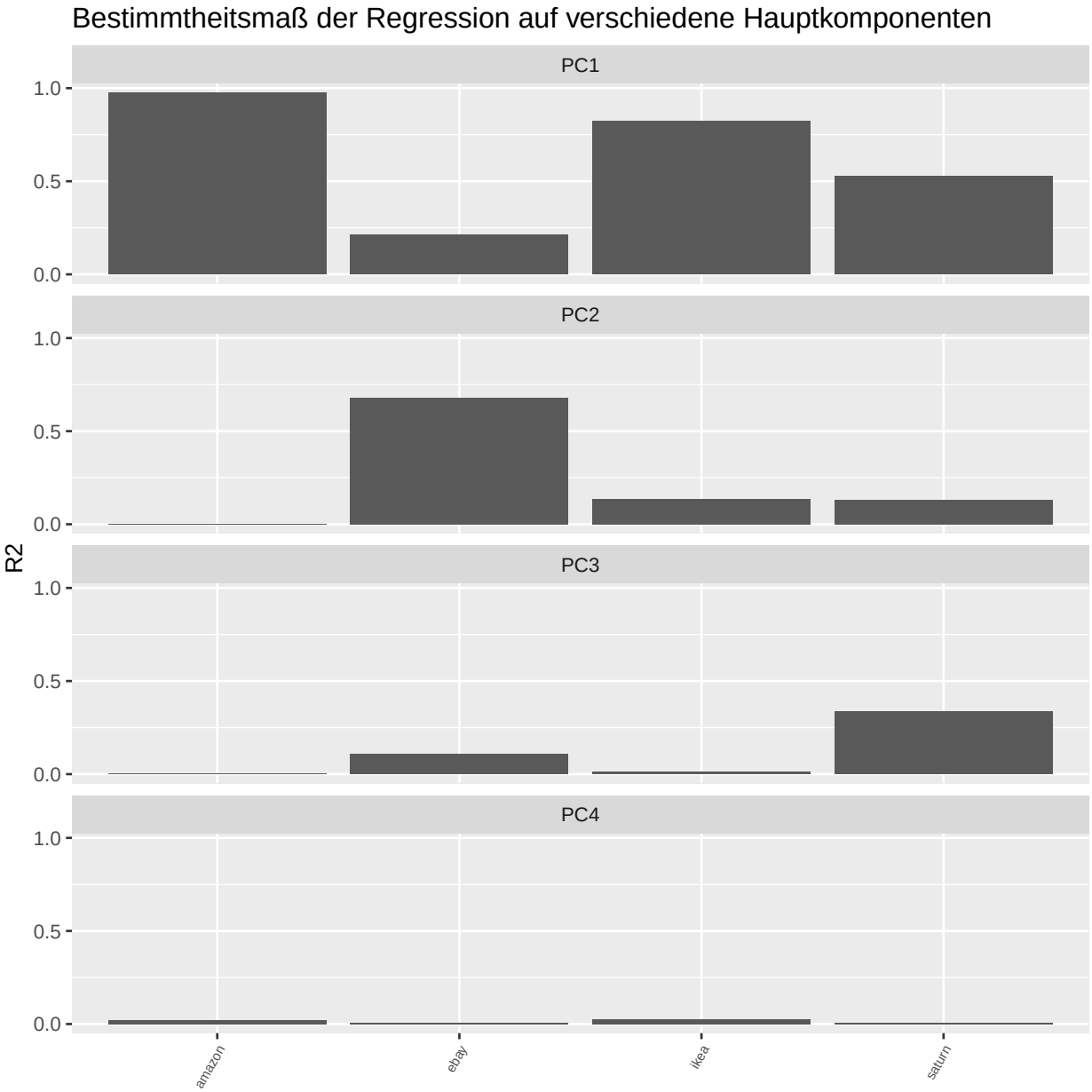
factorR2

Für schon berechnete Faktoren `factors` aus Zeitreihen `series` ist dies eine Methode, die Erklärungskraft der Faktoren zu bestimmen. Dabei wird für jeden Faktor eine Regression auf jede Zeitreihe vorgenommen und das jeweilige R^2 in einer Tabelle abgetragen. Mit Wahl des Parameters `plot=TRUE` wird zusätzlich ein Barplot ausgegeben.

```
dat <- pca(keywords = c("ikea", "saturn", "amazon", "ebay"),
    categories = 0,
    geo = "DE",
    start = "2006-01-01",
    end = Sys.Date(),
    components = max(length(keywords), length(categories)))
#> [time]: 'date'

series <- dat %>% select(date, 6:9)
factors <- dat %>% select(date, 2:5)

factorR2(series, factors, plot = T)
#> $res
#> # A tibble: 4 x 5
#>   factors   ikea  saturn   amazon    ebay
#>   <chr>    <dbl>   <dbl>    <dbl>   <dbl>
#> 1 PC1     0.821  0.528   0.977    0.211
#> 2 PC2     0.138  0.131   0.00113  0.678
#> 3 PC3     0.0140 0.335   0.000883 0.106
#> 4 PC4     0.0270 0.00702 0.0212   0.00523
#>
#> $plot
```


Bestimmtheitsmaß der Regression auf verschiedene Hauptkomponenten
PC1
PC2
PC3
PC4
R2
1.0
0.5
0.0
amazon
ebay
ikea
saturn