

Einlesen von Daten und weitere Regressionsbeispiele

1) Daten Einlesen und DAX-Composition:

a) Mit dem R-Befehl

```
read.table()
```

können Sie Daten von einem .txt oder .csv file ('csv' für 'comma seperated values') nach R importieren. Machen Sie sich mit diesem Befehl vertraut, indem Sie etwa mit

```
?read.table()
```

die R-Hilfe-Seiten aufrufen oder auf der Seite 64 im Kapitel 3.1.3 des R-Skriptes auf der Vorlesungshomepage nachschauen, unter

<http://hsrm-mathematik.de/WS201920/semester5/0ekonometrie/GerritEichner-UniGiessen-R1-4.pdf>

Importierte Daten werden in R typischerweise als ein `data.frame` angelegt. Über diesen Datentyp können Sie sich etwa in dem Kapitel 2.10 im R-Skript auf Seite 49 informieren. Sie sollten sich auf jeden Fall das Kapitel 2.10.1: Indizierung anschauen, damit Sie wissen, wie man etwa auf einzelne Daten-Elemente oder auf eine bestimmte Spalte eines Data Frames zugreifen kann. Laden Sie sich dann von der Vorlesungshomepage die Datei

```
DAXwithComp.txt
```

herunter und importieren Sie die Daten nach R, speichern Sie sie etwa in dem Dataframe `daxwithcomp`.

b) Wir wollen den Zeitraum von 2011 - 2014 betrachten, löschen Sie dazu alle Daten vor 2011. Mit dem Befehl `na.omit(daxwithcomp)` können Sie alle Zeilen mit NA's entfernen, von den 995 Zeilen bleiben dann noch 957 Zeilen übrig. Plotten Sie die Werte des DAX und, sagen wir, der Allianz- und der BASF-Aktie als Funktion von der Zeit. Dazu müssen Sie den DAX und die beiden Einzel-Aktien vorher aus den entsprechenden Spalten des `daxwithcomp`-DataFrames extrahieren.

c) Speichern Sie dann die Daten der 30 Einzelaktien in einer 957×30 Matrix `uls`. Bestimmen Sie jetzt die (approximative) Zusammensetzung des DAX, d.h. die Zahlen λ_i der Identität (nur approximativ, die Underlyings und die λ_i sind nicht über 4 Jahre lang dieselben)

$$\text{DAX}(t_k) = \sum_{i=1}^{30} \lambda_i \text{UL}_i(t_k)$$

mit Hilfe einer linearen Regression. Stellen Sie den DAX zusammen mit dem Regression-Fit in einem Diagramm dar.

2) Wachstumsrate der US-Wirtschaft aus S&P500-Daten:

Wir möchten den S&P500, den Aktienindex der 500 grössten US-amerikanischen Unternehmen, dazu benutzen, um eine konstante Wachstumsrate der amerikanischen Wirtschaft der letzten 65 Jahre, von 1950 bis 2015, zu berechnen. Dazu machen wir den Ansatz (das macht nur Sinn bei sehr grossen Zeiträumen)

$$\text{S\&P500}_t = S_{t_0} \times e^{r(t-t_0)} \quad (1)$$

- a) Laden Sie sich die Zeitreihe für den S&P500 von der Vorlesungshomepage herunter und importieren Sie die Daten nach R. Plotten Sie S\&P500_t als Funktion von t und plotten Sie ebenfalls $\log[\text{S\&P500}_t]$ als Funktion von t . Versuchen Sie, die Plots so hinzubekommen, dass auf der horizontalen Achse, der t -Achse, das tatsächliche Datum angezeigt wird. Sie brauchen dazu ein paar Informationen zum Formatieren von Datütern, die Sie etwa in `Date-Formatting-in-R.pdf` auf der Vorlesungshomepage finden können (das ist ein ziemliches Gefrickel und deshalb **nicht** Klausur-relevant).
- b) Bestimmen Sie jetzt die Wachstumsrate r und die Konstante S_{t_0} mit Hilfe einer linearen Regression. Logarithmieren Sie dazu die Gleichung (1):

$$\log[\text{S\&P500}_t] = \log S_{t_0} + r(t - t_0)$$

Dann können wir also die Wachstumsrate r als Regressionskoeffizient β_1 des Regressionsproblems

$$y_t = \beta_0 + \beta_1 x_t + \text{error}$$

bekommen mit $y_t = \log[\text{S\&P500}_t]$ und $x_t = t - t_0$. Zeigen Sie jeweils den Regression-Fit zusammen mit den beiden Plots von S\&P500_t und $\log[\text{S\&P500}_t]$ aus Teil (a).