

Matthias Gehrke

# **Angewandte empirische Methoden in Finance & Accounting**

---

Umsetzung mit R

**DE GRUYTER**  
OLDENBOURG

# Inhalt

Vorwort — VII

Abbildungsverzeichnis — XV

Tabellenverzeichnis — XVII

Verzeichnis der R-Codes — XIX

Verzeichnis der R-Grafiken — XXV

- 1 Einführung — 1**
  - 1.1 Über dieses Buch — 1
  - 1.2 Durchführung eines empirischen Projekts — 1
    - 1.2.1 Vorbereitung und Datenerhebung — 1
    - 1.2.2 *Explorative Datenanalyse* — 1
    - 1.2.3 Modellierung und Ergebnisdarstellung — 1
  - 1.3 Grundlagen statistischer Modellierung — 2
  - 1.4 Inferenz — 4
  - 1.5 Hinweise zu R — 4
  - 1.6 Literatur — 6
  
- 2 Lineare Regression und Regressionsdiagnostik — 7**
  - 2.1 Grundlagen der linearen Regression — 7
  - 2.2 Verwendete R-Pakete — 10
  - 2.3 Einführungsbeispiel — 10
  - 2.4 Interpretation und Inferenz der Koeffizienten — 15
    - 2.4.1 Interpretation der Koeffizienten – metrisch — 15
    - 2.4.2 Interpretation der Koeffizienten – kategorial — 19
    - 2.4.3 Inferenz der Koeffizienten — 21
  - 2.5 Erweiterung der Modellgleichung — 23
    - 2.5.1 Interaktion — 23
    - 2.5.2 Verschachtelte Modelle — 28
    - 2.5.3 Formelsyntax — 31
  - 2.6 Globale Modellgüte und Inferenz — 31
    - 2.6.1 Globale Modellgüte — 31
    - 2.6.2 Inferenz für das Modell als Ganzes — 36
  - 2.7 Anwendungsvoraussetzungen für die lineare Regression — 36
    - 2.7.1 Variablentransformation — 38
    - 2.7.2 Robustheit und einflussreiche Beobachtungen — 42

2.7.3	Nichtlinearer Zusammenhang und Fehlspezifikation —	48
2.7.4	Erwartungswert der Residuen —	57
2.7.5	Korrelation zwischen Residuen und unabhängigen Variablen —	59
2.7.6	Heteroskedastizität —	60
2.7.7	Autokorrelation —	68
2.7.8	Multikollinearität —	76
2.7.9	Normalverteilung der Residuen —	80
2.7.10	Verschiedene diagnostische Grafiken —	84
2.8	Modellselektion und Wichtigkeit der Variablen —	86
2.8.1	Modellselektion —	86
2.8.2	Erklärung —	86
2.8.3	Modellselektion mit der <code>step()</code> -Funktion —	87
2.8.4	Wichtigkeit der Variablen —	89
2.8.5	Vorhersage —	90
2.9	Literatur —	92
2.9.1	Weiterführende Literatur —	92
2.9.2	Anwendungsbeispiele —	93
<b>3</b>	<b>Panelregression —</b>	<b>94</b>
3.1	Grundlagen der Panelregression —	94
3.2	Verwendete R-Pakete —	95
3.3	Vorbereitung der Panelregression —	96
3.4	Ergänzungen für die explorative Datenanalyse —	97
3.5	Pooled Model —	98
3.5.1	Durchführung der Pooled Regression —	99
3.5.2	Interpretation des Pooled Model —	100
3.6	Fixed Effects Model —	101
3.6.1	Anwendung des Fixed Effects Model —	103
3.6.2	Überprüfung auf fixe Effekte —	104
3.6.3	Vor- und Nachteile des Fixed Effects Model —	105
3.7	Random Effects Model —	105
3.7.1	Anwendung des Random Effects Model —	106
3.7.2	Überprüfung auf zufällige Effekte —	108
3.7.3	Überprüfung der Voraussetzungen für das Random Effects Model —	108
3.7.4	Vor- und Nachteile des Random Effects Model —	109
3.8	Auswahl des Modells —	109
3.9	Regressionsdiagnostik —	111
3.9.1	Heteroskedastizität —	111
3.9.2	Autokorrelation —	113
3.9.3	Querschnittskorrelation —	115
3.9.4	Multikollinearität —	117
3.10	Literatur —	118

- 3.10.1 Weiterführende Literatur — 118
- 3.10.2 Anwendungsbeispiele — 118
  
- 4 Logistische Regression — 119**
  - 4.1 Grundlagen der logistischen Regression — 119
    - 4.1.1 Transformation — 119
    - 4.1.2 Modellgleichung — 121
    - 4.1.3 Linkfunktion — 121
    - 4.1.4 Maximum-Likelihood-Schätzung — 122
  - 4.2 Verwendete R-Pakete — 123
  - 4.3 Interpretation der Koeffizienten — 124
    - 4.3.1 Direkte Interpretation — 124
    - 4.3.2 Interpretation über das Chancenverhältnis — 125
    - 4.3.3 Interpretation über den marginalen Effekt — 130
  - 4.4 Globale Modellgüte — 134
    - 4.4.1 Devianz — 134
    - 4.4.2 Likelihood-Quotienten-Test — 136
    - 4.4.3 Pseudo- $R^2$  — 137
  - 4.5 Inferenz der Koeffizienten — 137
    - 4.5.1 Wald-Test — 137
    - 4.5.2 Likelihood-Quotienten-Test — 138
    - 4.5.3 Inferenz der marginalen Effekte — 139
  - 4.6 Interaktion — 141
  - 4.7 Regressionsdiagnostik — 146
    - 4.7.1 Einflussreiche Beobachtungen — 146
    - 4.7.2 Nichtlinearer Zusammenhang — 150
    - 4.7.3 Multikollinearität — 152
    - 4.7.4 Robuste Varianz-Kovarianz-Schätzer — 153
  - 4.8 Modellselektion und Wichtigkeit der Variablen — 154
    - 4.8.1 Modellselektion — 154
    - 4.8.2 Wichtigkeit der Variablen — 156
  - 4.9 Klassifikationseigenschaften — 156
    - 4.9.1 Konfusionsmatrix — 156
    - 4.9.2 ROC-Kurve und AUC-Wert — 160
    - 4.9.3 Liftwerte und Liftkurve — 163
  - 4.10 Literatur — 166
    - 4.10.1 Weiterführende Literatur — 166
    - 4.10.2 Anwendungsbeispiele — 166
  
- 5 Klassifikation und Regression mit Bäumen und Random Forest — 167**
  - 5.1 Grundlagen baumbasierter Verfahren — 167
  - 5.2 Verwendete R-Pakete — 168

5.3	Klassifikations- und Regressionsbäume — 169
5.3.1	Klassifikationsbäume — 169
5.3.2	Wichtige Parameter der Funktion <code>rpart()</code> — 176
5.3.3	Beschneiden der Bäume — 177
5.3.4	Regressionsbäume — 180
5.3.5	Vor- und Nachteile baumbasierter Verfahren — 187
5.4	Random Forest — 188
5.4.1	Bootstrapping — 188
5.4.2	Bootstrapping bei Random Forest — 189
5.4.3	Random Forest für die Klassifikation — 189
5.4.4	Random Forest für die Regression — 194
5.4.5	Vorhersage — 196
5.4.6	Wichtige Parameter von <code>randomForest()</code> und <code>cforest()</code> — 197
5.4.7	Wichtigkeit der Variablen — 197
5.4.8	Weitere Möglichkeiten — 201
5.4.9	Vor- und Nachteile von Random Forest — 201
5.5	Literatur — 202
5.5.1	Weiterführende Literatur — 202
5.5.2	Anwendungsbeispiele — 202
<b>6</b>	<b>Hauptkomponentenanalyse — 203</b>
6.1	Grundlagen — 203
6.2	Verwendete R-Pakete — 207
6.3	Beispiel einer Hauptkomponentenanalyse — 208
6.3.1	Eignung der Daten — 209
6.3.2	Anzahl der Hauptkomponenten — 211
6.3.3	Durchführung und Interpretation der Hauptkomponentenanalyse — 213
6.3.4	Zusammenfassung der Hauptkomponenten — 217
6.4	Literatur — 219
6.4.1	Weiterführende Literatur — 219
6.4.2	Anwendungsbeispiele — 220
<b>7</b>	<b>Analyse von Zeitreihen — 221</b>
7.1	Einführung — 221
7.2	Verwendete R-Pakete — 221
7.3	Grundbegriffe — 222
7.3.1	Stationarität — 223
7.3.2	Autokovarianz- und Autokorrelationsfunktion — 223
7.3.3	Stochastischer Prozess — 226
7.3.4	Differenzbildung — 227
7.3.5	Integration und Kointegration — 228
7.3.6	Stochastischer und deterministischer Trend — 231

- 7.3.7 Tests auf Stationarität — 232
- 7.4 AR- und MA-Prozesse — 236
  - 7.4.1 AR-Prozess — 236
  - 7.4.2 MA-Prozess — 238
- 7.5 AR(I)MA-Modelle — 240
  - 7.5.1 Residuenanalyse — 240
  - 7.5.2 Beispiel für ein ARIMA-Modell — 241
  - 7.5.3 Verschiedene diagnostische Plots — 246
  - 7.5.4 Automatische Schätzung der Modellparameter — 248
  - 7.5.5 Prognose — 248
- 7.6 Rendite und Volatilität — 250
  - 7.6.1 Rendite — 250
  - 7.6.2 Volatilität — 250
  - 7.6.3 Gewichtung der früheren Varianzen — 251
  - 7.6.4 EWMA-Modelle — 252
- 7.7 (G)ARCH-Modelle — 252
  - 7.7.1 ARCH-Modelle — 253
  - 7.7.2 GARCH-Modelle — 256
  - 7.7.3 Beispiel eines GARCH-Modells — 258
  - 7.7.4 Ausblick — 267
- 7.8 Literatur — 268
  - 7.8.1 Weiterführende Literatur — 268
  - 7.8.2 Anwendungsbeispiele — 268
- 8 Ereignisstudie — 269**
  - 8.1 Einführung — 269
  - 8.2 Renditen — 269
    - 8.2.1 Normale Rendite — 270
    - 8.2.2 Ereignis- und Schätzfenster — 273
    - 8.2.3 Abnormale Renditen — 274
  - 8.3 Testverfahren — 275
    - 8.3.1 t-Test — 275
    - 8.3.2 Cross-Sectional-Dependence-Test (CSD-Test) — 276
    - 8.3.3 Cross-Sectional-Independence-Test (CSI-Test) — 276
    - 8.3.4 Corrado-Rang-Test — 277
  - 8.4 Verwendete R-Pakete — 278
  - 8.5 Durchführung einer Ereignisstudie — 278
  - 8.6 Ausblick — 284
  - 8.7 Literatur — 284
    - 8.7.1 Weiterführende Literatur — 284
    - 8.7.2 Anwendungsbeispiele — 285

**XIV — Inhalt**

**A Anhang — 287**

A.1 Verwendete R-Pakete — **287**

A.2 Verwendete Datensätze — **290**

**Quellennachweise — 293**

**Literatur — 295**

**Stichwortverzeichnis — 303**

**Verzeichnis der verwendeten R-Funktionen — 311**