Aufgabenblatt

Die Aufgaben können in **2er Gruppen** bearbeitet werden. Für jede Aufgabe sollen Sie ein R Skript erstellen, das Sie als „`Aufgabennr.Vorname1.Vorname2.R`“ speichern, also z.B. „`3.2.Bernd.Ute.R`“ für das R Skript zur Aufgabe 3.2.
Abgabe: **18.02.2010 bis 18:00 Uhr** per Email. Besprechung: **19.02.2010 um 13:30** im SRA.

7 ARMA-Modelle: Theorie

Aufgabe 7.1. (5 Punkte)
Rechnen Sie (Pen & Paper) nach, dass für einen MA(q)-Prozess

$$X_t = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q}$$

mit $(W_t) \sim \mathrm{WN}(0, \sigma^2)$ gilt

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^{q} \theta_i^2} & h = 1, \dots, q \\ 0 & h > q \end{cases} \tag{1}$$

wobei $\theta_0 = 1$.

Aufgabe 7.2. (5 Punkte)
Bringen Sie die folgenden Modelle in ARMA(p,q)-Darstellung mit Polynomen Φ und Θ. Überprüfen Sie, ob die Modelle invertierbar sind. Überprüfen Sie, ob bereits ein ARMA(p-1, q-1)-Modell genügt.

(a) $X_t = X_{t-1} - \frac{1}{4} X_{t-2} + W_t + \frac{1}{2} W_{t-1}$

(b) $X_t = \frac{7}{10} X_{t-1} - \frac{1}{10} X_{t-2} + W_t - \frac{3}{2} W_{t-1}$

(c) $X_t = \frac{3}{2} X_{t-1} - \frac{1}{2} X_{t-2} + W_t - \frac{1}{3} W_{t-1} + \frac{1}{6} W_{t-2}$

8 Modellierung mittels ARMA-Prozessen

Aufgabe 8.1. (10 Punkte)
Lesen Sie mit dem Befehl `source` die Datei `wechselkurs.R` ein. Sie finden dort die Zeitreihe `wechselkurs`. Diese enthält die monatlichen Durchschnitte des Wechselkurses von Euro in Dollar.

(a) Über welchen Zeitraum wurden diese Daten erhoben?

(b) Zentrieren Sie die Daten, und arbeiten Sie im Folgenden mit der zentrierten Zeitreihe `wk.zentr`.

(c) Entscheiden Sie anhand der Box-Jenkins-Methode, welches ARMA(p,q)-Modell mit $p+q \leq 2$ die Daten am besten beschreibt. Fitten Sie das entsprechende Modell an die Daten (beachten Sie, dass Sie mit zentrierten Daten arbeiten), und überprüfen Sie anhand der Residuen ihre Hypothese.

(d) Schreiben Sie eine Funktion `best.order(x)`, die für eine gegebene Zeitreihe x anhand des AIC das best geeigneteste ARMA(p,q)-Modell mit $p+q \leq 2$ auswählt. *Hinweis*: Auf das AIC des gefitteten Modells können Sie mittels `arima(...)$aic` zugreifen.

(e) Machen Sie mit dem in (d) gewählten Modell eine 1-Schritt-Vorhersage für den Wechselkurs (denken Sie an Ihre bisherigen Transformationen!).

Aufgabe 8.2. (10 Punkte)
Lesen Sie mit dem Befehl `source` die Datei `temperaturen.R` ein, und betrachten Sie die Zeitreihe `Global.annual`.

(a) Berechnen Sie die Zeitreihe der Differenzen zum Vorjahreswert, und arbeiten Sie im Folgenden mit dieser Zeitreihe `temp`.

(b) Entscheiden Sie anhand der Box-Jenkins-Methode, welches ARMA(p,q)-Modell mit $p+q \leq 2$ die Daten am besten beschreibt. Fitten Sie das entsprechende Modell an die Daten (berechnen Sie auch einen `intercept`-Term), und überprüfen Sie anhand der Residuen ihre Hypothese.

(c) Nutzen Sie wieder die selbstgeschriebene Funktion `best.order(x)`, um anhand des AIC das best geeigneteste ARMA(p,q)-Modell mit $p+q \leq 2$ auszuwählen.

(d) Machen Sie anhand dieses Modells eine 1-Schritt-Prognose für die globale Durchschnittstemperatur (denken Sie an die bisher durchgeführten Transformationen!).

Aufgabe 8.3. (12 Punkte)
Lesen Sie mit dem Befehl `source` die Datei `EnergieverbrauchUSA.R` ein. Sie finden dort die Variable `enUSA` mit dem monatlichen Gesamtenergieverbrauch der Vereinigten Staaten in Exajoule (10^{18} Joule) von 1973 bis 2008.

Plotten Sie die Zeitreihe. Liegen saisonale Schwankungen vor? Warum?

Wählen Sie nun einen Zeitbereich, in welchem ein linearer Trend vorliegen könnte. Ziel dieser Aufgabe ist es, eine Modellprognose mit den realen Daten zu vergleichen. Passen Sie dazu ein Modell an die ersten Beobachtungen in diesem Zeitbereich an, um daraus Vorhersagen für die letzten drei Jahre des gewählten Zeitbereiches zu machen. Gehen Sie dazu wie folgt vor:

(a) Schränken Sie die Zeitreihe auf den ausgewählten Zeitbereich (ohne dessen letzte drei Jahre) ein, und weisen Sie diese der Variablen `Energie.part` zu.

(b) Berechnen Sie (für `Energie.part`) einen linearen Trend und additive Saisonkomponenten.

(c) Nutzen Sie wieder die selbstgeschriebene Funktion `best.order(x)`, um für die Residuen des linearen Modells das best geeigneteste ARMA(p, q)-Modell mit $p + q \leq 2$ auszuwählen.

(d) Berechnen Sie eine Prognose für die ausgenommenen 3 Jahre (die Einzelprognosen beider Modelle werden additiv überlagert); und plotten Sie Prognose und tatsächlichen Verlauf in eine Grafik.

9 Simulation

Aufgabe 9.1. (8 Punkte)
Überprüfen Sie die Aussagen über ACF und PACF empirisch; indem Sie jeweils 3 Realisationen der Länge n folgender Prozesse erzeugen, (schreiben Sie eine geeignete Funktion in R dafür, wenn Sie noch Zeit haben; ansonsten gibt es `arima.sim`), und deren empirische ACF und PACF anzeigen lassen. Setzen Sie $X_{-1} = X_0 = 0$ bei AR-Prozessen.

(a) MA(1), $\theta_1 = 0.7$, $n = 100$.

(b) AR(1), $\phi_1 = 0.8$, $n = 100, 1000$

(c) AR(2), $\phi_1 = 0.1$, $\phi_2 = -0.8$ $n = 100, 10000$

(d) MA(4), $\theta_1 = 0.8$, $\theta_2 = 0.6$, $\theta_3 = 0.5$, $\theta_4 = 0.35$, $n = 1000$

(e) ARMA(1,1), $\phi_1 = 0.7$, $\theta_1 = 0.7$, $n = 1000$

Hinweis: Mit dem Befehl `par(mfrow=c(3,3))` zeigen Sie 9 Plots in einer Grafik an.

9.1 Zentrieren: ja oder nein

Um vernünftige Saisonfiguren mittels `lm` zu erhalten, war es ratsam, Zeit und Zeitreihenwerte zu zentrieren.

Tatsächlich ist dies für Prognosen in R aber eher hinderlich, da viele Terme mitgeschleppt werden müssen. Deshalb sollten Sie ab jetzt im Allgemeinen keine Zentrierungen mehr durchführen. Lineare Modelle sollten Sie in der Form

verb- lm(Zeitreihe time(Zeitreihe) +factor(cycle(Zeitreihe))) -

aufrufen. Die Saisonkomponenten können nun sehr groß ausfallen, außerdem fehlt die erste. Sie können diese allerdings leicht berechnen, da die Summe der Saisonkomponenten Null ergeben muss. Schreiben Sie `Intercept` und alle berechneten Saisonkomponenten in einen Vektor, und zentrieren Sie diesen. Das arithmetische Mittel ist nun die tatsächliche Verschiebung, und der zentrierte Vektor enthält nun vernünftige Saisonkomponenten, die durch die Saison hervorgerufen werden.

9.2 Vorhersagen aus linearen Modellen

Diese geschehen mit der Funktion `predict`, allerdings funktioniert der Parameter `n.ahead` bei linearen Modellen (z.B. Regressionsmodelle zur Bestimmung eines linearen Trends) nicht. Stattdessen muss `newdata` angegeben werden, und zwar in subtiler Form. Diese lässt sich anhand eines Beispiels am besten nachvollziehen. Sei `x` eine Zeitreihe. Die Variablenzuweisungen zu Beginn sind hier nötig.

```
ZEIT<-time(x); SAISON<-factor(cycle(x))
modell<-lm(x ~ ZEIT + SAISON)
y<-ts(start=c(...),end=c(...),f=12)
Bereich<-data.frame(ZEIT=time(y), SAISON=factor(cycle(y)))
y.dat<-predict(modell,newdata=Bereich)
y<-ts(y.dat,start=c(...),end=c(...),f=12)
```

Die Zeitreihe y enthält die vorhergesagten Werte, bei Start, Ende und Frequenz sind die entsprechenden Daten einzutragen, typischerweise wird der Beginn von y nach dem Ende von x liegen, und die Frequenz muss gleich bleiben.