**Regressionsmodelle mit Anwendungen in der Versicherungs- und Finanzwirtschaft**

**Probeklausur** **Wintersemester 2017/2018** **06.12.2018**

BITTE LESERLICH IN **DRUCKBUCHSTABEN** AUSFÜLLEN

Nachname: ..............................................................

Vornäme: ..............................................................

Matrikelnummer: | | | | | | | |

Studienfach: ..............................................................

**Unterschrift der/des Studierenden:**

Überprüfen Sie die Klausur auf Vollständigkeit, sie besteht aus **??** Seiten.

**Bemerkungen:**

| **Aufgabe** | **max. Pkt.** | **err. Pkt.** |
|---|---|---|
| 1 | 20 | |
| 2 | 20 | |
| 3 | 5 | |
| 4 | 15 | |
| 5 | 5 | |
| 6 | 10 | |
| 7 | 5 | |
| 8 | 12 | |
| 9 | 8 | |
| **Summe** | **100** | |
| **Bonus** | **6** | |
| **Note** | | |

## Aufgabe 1: Multiple Choice (20 Punkte)

Kreuzen Sie an, welche der nachfolgenden Aussagen wahr bzw. falsch sind. Beachten Sie:

Jede Aussage ist entweder wahr oder falsch.
Richtig gesetztes Kreuz = 2 Punkte
Kein gesetztes Kreuz = 1 Punkt
Falsch gesetztes Kreuz = 0 Punkte

| | | wahr | falsch |
|---|---|---|---|
| (a) | Bei linksschiefen Verteilungen nimmt das dritte zentrierte theoretische Moment besonders große Werte an. | ☐ | ☐ |
| (b) | Die Power-Transformation wird verwendet um rechtsschiefe Verteilungen so zu transformieren, damit sie annähernd symmetrisch sind. | ☐ | ☐ |
| (c) | Sei $\mathbf{X} = (X_1, X_2, \ldots, X_n) \sim \mathbb{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ und $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ eine Zerlegung von $\mathbf{X}$ in zwei Zufallsvektoren, dann ist $X_1 \vert X_2$ ebenfalls multivariat Normalverteilt. | ☐ | ☐ |
| (d) | Das adjustierte Bestimmtheitsmaß $R_a^2$ kann größere Werte annehmen als das Bestimmtheitsmaß $R^2$. | ☐ | ☐ |
| (e) | Das Lognormal-Verfahren ist ein auf dem verallgemeinerten-linearen Modell basierendes Verfahren. | ☐ | ☐ |
| (f) | Jeder schwach stationäre Prozess weist eine über die Zeit konstante Varianz auf. | ☐ | ☐ |
| (g) | Der Lag-Operator verschiebt den Zeitindex des stochastischen Prozesses um eine Einheit in die Zukunft. | ☐ | ☐ |
| (h) | Ein White-Noise Prozess ist ein Spezialfall aus der Klasse der MA-Prozesse. | ☐ | ☐ |
| (i) | Der Lagrange-Multiplikator-Test prüft bei einer vorgegebenen Zeitreihe das Vorhandensein von Heteroskedastizität. | ☐ | ☐ |
| (j) | Bei der Parameterschätzung von AR($p$)-Prozessen liefern die Kleinste-Quadrate-, Momenten- und Maximum-Likelihood-Methode immer die selben Schätzungen. | ☐ | ☐ |

## Aufgabe 2: Wissensfragen (20 Punkte)

a) Was versteht man unter einer Schadenreserve? (2 Punkte)

b) Nennen Sie die vier zentralen Annahmen, die dem Klassischen linearen Modell zugrunde liegen. (2 Punkte)

c) Was versteht man unter dem Begriff heteroskedastischer Fehler? Nennen Sie zusätzlich einen Datentypen bei dem häufig heteroskedastische Fehler auftreten? (2 Punkte)

d) Nennen Sie zwei Kennzahlen mit deren Hilfe man beim Klassischen linearen Modell beurteilen kann, ob es sich bei einer Beobachtung um einen Ausreißer handelt. (2 Punkte)

e) Was versteht man unter einem saturierten Modell und was ist das besondere an den Schätzungen $\widehat{y}_i$ in diesem Modell? (2 Punkte)

f) Was würde mit dem KQ-Schätzer passieren, wenn man bei einer kategorialen Variablen keine Referenzklasse bilden würde und das Modell für jede Klasse eine Binäre Variable enthält. (2 Punkte)

g) Was ist ein stochastischer Prozess? (2 Punkte)

h) Nennen Sie eine Erweiterung für das GARCH-Modell. (2 Punkte)

i) Nennen Sie einen Test mit dem überprüft werden kann, ob ein ARMA-Prozess nicht schwach stationär ist. (2 Punkte)

j) Welche Bedingungen muss ein stochastischer Prozess $(\varepsilon_t)_{t \in \mathbb{Z}}$ erfüllen, damit es sich bei dem Prozess um einen White-Noise Prozess handelt? (2 Punkte)

## Aufgabe 3: Multivariate Normalverteilung (5 Punkte)

Der Zufallsvektor $\mathbf{X} = (X_1, X_2, X_3)^T$ sei dreidimensional $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-verteilt mit

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix} \quad \text{und} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 2 & -1 \\ 2 & 25 & -10 \\ -1 & -10 & 5 \end{pmatrix}$$

Bestimmen Sie die Verteilung von $X_2$ gegeben $x_1 = -2$ und $x_3 = 7$.

---

## Aufgabe 4: Residuenanalyse (15 Punkte)

Es wird das Klassische lineare Modell der Form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

mit $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{E})$ betrachtet. Für den Beobachtungsvektor $\mathbf{y}$ und für die Designmatrix $\mathbf{X}$ hat man

$$\mathbf{y} = \begin{pmatrix} 6 \\ 10 \\ 12 \\ 30 \end{pmatrix} \quad \text{und} \quad \mathbf{X} = \begin{pmatrix} 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 10 \end{pmatrix}$$

erhalten. Mit $\mathbf{y}$ und $\mathbf{X}$ erhält man für den Kleinste-Quadrate-Schätzer

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta}_0 & \widehat{\beta}_1 \end{pmatrix}^T = \begin{pmatrix} -\frac{124}{29} & \frac{99}{29} \end{pmatrix}^T$$

und für die hat-Matrix

$$\mathbf{H} = \frac{1}{116} \begin{pmatrix} 54 & 44 & 34 & -16 \\ 44 & 38 & 32 & 2 \\ 34 & 32 & 30 & 20 \\ -16 & 2 & 20 & 110 \end{pmatrix}.$$

a) Berechnen Sie das standardisierte Residuum $r_4$ und beurteilen Sie anhand einer gängigen Daumenregel, ob es sich um einen Ausreißer handelt. (6 Punkte)

b) Bestimmen Sie das studentisierte Residuum $r_4^*$. und beurteilen Sie anhand einer gängigen Daumenregel, ob es sich bei der Beobachtung $y_4$ um einen Ausreißer handelt. (9 Punkte)

---

## Aufgabe 5: Modellüberprüfung und Modellwahl (5 Punkte)

Gegeben ist ein Datensatz mit der Designmatrix

$$X = \begin{pmatrix} 1 & 4 \\ 1 & 5 \\ 1 & 10 \end{pmatrix}.$$

Bestimmen Sie für diesen Datensatz die Hat-Matrix $H$. Ist eine der Beobachtungen selbst-schätzend? Verwenden Sie hierfür eine Daumenregel und beurteilen sie, ob die Anwendung dieser Daumenregel in diesem Fall sinnvoll ist.
Hinweis:

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \frac{1}{62}\begin{pmatrix} 141 & -19 \\ -19 & 3 \end{pmatrix}$$

---

## Aufgabe 6: Verallgemeinerte lineare Modelle (10 Punkte)

Die Finanzabteilung eines großen Unternehmens verwendet ein Verallgemeinertes lineares Modell mit $k = 3$ erklärenden Variablen. Einige der Experten schlagen eine Erweiterung des bisherigen Modells um 4 zusätzliche Variablen vor. Zum Vergleich der beiden Modelle wurde eine Stichprobe von $n = 12$ Beobachtungen erhoben. Für das kleinere Modell $M_1$ hat man einen Log-Likelihood Wert von

$$l\left(\widehat{\boldsymbol{\beta}}_{M_1;y_1,\ldots,y_n}\right) = -10{,}6$$

und für das größere Modell $M_2$ einen Wert von

$$l\left(\widehat{\boldsymbol{\beta}}_{M_2;y_1,\ldots,y_n}\right) = -6{,}4$$

erhalten. Der Log-Likelihood Wert des saturierten Modells beträgt

$$l\left(\widehat{\boldsymbol{\beta}}_{S;y_1,\ldots,y_n}\right) = -2.$$

a) Prüfen Sie bei einem Signifikanzniveau von $\alpha = 0{,}1$, ob die Aufnahme der zusätzlichen Variablen eine signifikante Verbesserung liefert. (3 Punkte)

b) Für welches der beiden Modelle würde sich das Unternehmen entscheiden, wenn es das Informationskriterium von Akaike(AIC) als Entscheidungskonzept heranziehen würde? (2 Punkte)

c) Für welches der beiden Modelle würde sich das Unternehmen entscheiden, wenn es das Bayesianische Informationskriterium(BIC) als Entscheidungskonzept heranziehen würde? (2 Punkte)

d) Angenommen, das Unternehmen überlegt weder $M_1$ noch $M_2$ zu verwenden und stattdessen das saturierte Modell S. Für welches der drei Modelle sollte sich das Unternehmen entscheiden, wenn das Informationskriterium von Akaike(AIC) bzw. das Bayesianische Informationskriterium(BIC) verwendet wird? (3 Punkte)

---

## Aufgabe 7: Schwache Stationarität (5 Punkte)

Es seien $(X_t)_{t\in\mathbb{Z}}$ und $(Y_t)_{t\in\mathbb{Z}}$ zwei schwach stationäre und stochastisch unabhängige Prozesse mit den Autokovarianzfunktionen $\gamma_x(h)$ und $\gamma_y(h)$ und den Erwartungswerten $\mu_x$ und $\mu_y$. Untersuchen Sie den Prozess $(Z_t)_{t\in\mathbb{Z}}$ mit

$$Z_t = X_t Y_t$$

auf schwache Stationarität.

---

## Aufgabe 8: ARMA-Prozesse (12 Punkte)

Betrachtet wird der stochastische Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = X_{t-1} - 1{,}25X_{t-2} + \varepsilon_t - 0{,}25\varepsilon_{t-2}$$

mit $\varepsilon_t \sim WN(0,\sigma^2)$.

a) Um was für einen Prozess handelt es sich bei $(X_t)_{t\in\mathbb{Z}}$? (1 Punkt)

b) Untersuchen Sie den Prozess auf Kausalität und Invertierbarkeit. (5 Punkte)

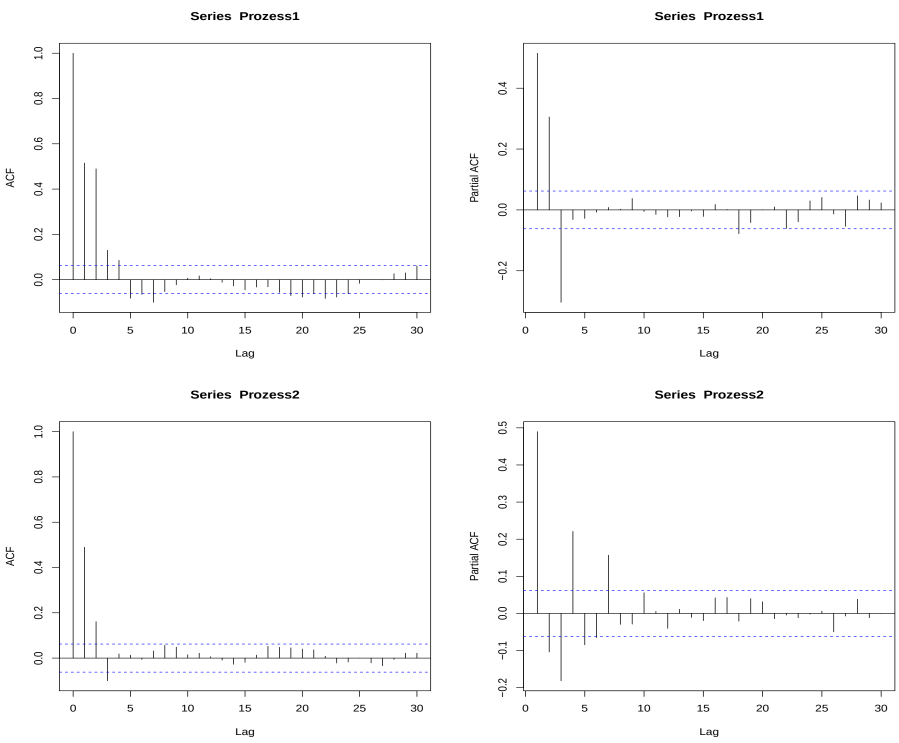
Betrachten Sie nun die stochastischen Prozesse $(W_t)_{t\in\mathbb{Z}}$ und $(Z_t)_{t\in\mathbb{Z}}$ mit

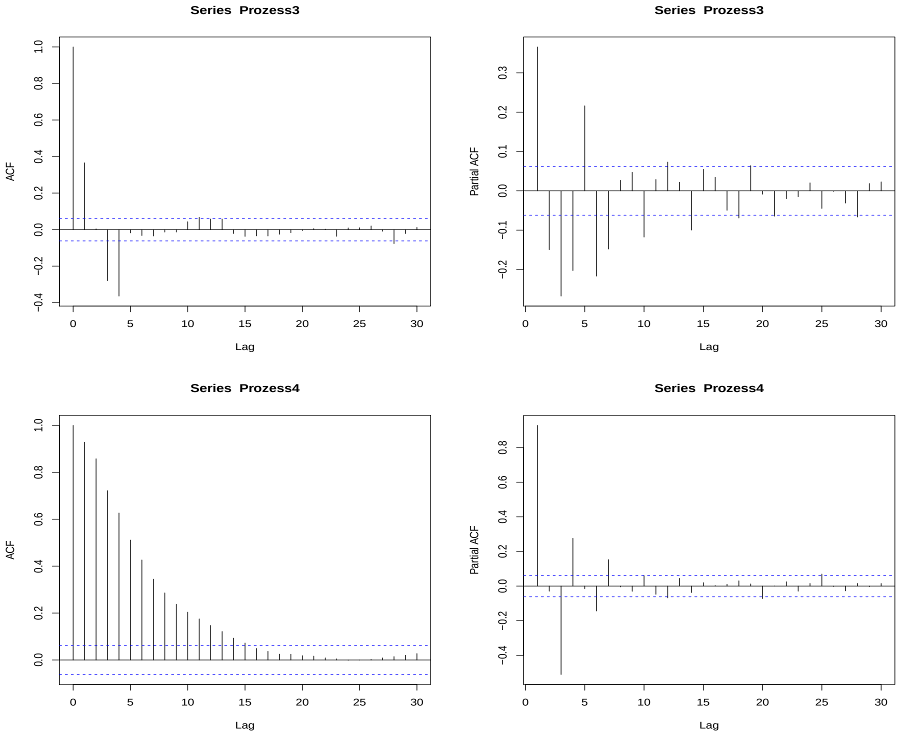
$$\begin{aligned} W_t &= \varepsilon_t + 0{,}8\varepsilon_{t-1} - 0{,}8\varepsilon_{t-4}\ , \\ Z_t &= 0{,}45Z_{t-1} + 0{,}4Z_{t-2} - 0{,}3Z_{t-3} + \varepsilon_t \end{aligned}$$

für $t \in \mathbb{Z}$ und $\varepsilon_t \sim WN(0,\sigma^2)$.

c) Ordnen Sie den Prozessen $(W_t)_{t\in\mathbb{Z}}$ und $(Z_t)_{t\in\mathbb{Z}}$ die passenden empirischen Autokorrelationsfunktionen und empirischen partiellen Autokorrelationsfunktionen zu und begründen Sie ihre Zuordnungen. (3 Punkte)

d) Geben Sie zu den zwei nicht zugeordneten Prozessen die jeweils passende Prozessklasse an. (3 Punkte)




Series Prozess3
ACF
Lag
Series Prozess3
Partial ACF
Lag
Series Prozess4
ACF
Lag
Series Prozess4
Partial ACF
Lag

## Aufgabe 9: Schätzung und Prognose (8 Punkte)

Betrachtet wird der stochastische Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = \varepsilon_t + 0{,}3\varepsilon_{t-1} - 0{,}5\varepsilon_{t-2}$$

wobei $\varepsilon \sim WN(0, \sigma^2)$ gilt.

a) Geben Sie die Autokorrelationsfunktion $\rho(h)$ des stochastischen Prozesses $(X_t)_{t\in\mathbb{Z}}$ an. (3 Punkte)

b) Bestimmen Sie den besten affin-linearen Prädiktor

$$\widehat{X}_{T+h|T}^{opt} = a_0 + \sum_{t=1}^{T} a_{T+1-t} X_t$$

für $T = 2$ und $h = 2$. (5 Punkte)

---