

Aufgabe 1 (18 Punkte)

Gegeben sei das folgende lineare Regressionsmodell:

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_N), \quad (1)$$

wobei X eine deterministische $(N \times K)$ -Matrix ist mit $\text{Rang}(X) = K$.

- a) (2 Punkte) Welche Modellannahme ist im Fall einer exakten Multikollinearität verletzt?
- b) (3 Punkte) Nehmen Sie an, Gleichung (1) stelle das wahre Modell dar, während für die KQ-Schätzung die folgende Spezifikation verwendet wird: $y = X\beta + Z\gamma + u$.
Nennen Sie die Auswirkungen einer solchen Überspezifikation bzgl. der Erwartungstreue und Effizienz des KQ-Schätzers $\hat{\beta}_{KQ}$.
Weshalb würde eine Unterspezifikation ein größeres Problem darstellen?
- c) (4 Punkte) Tabelle 1 zeigt das Ergebnis des White-Tests für die Residuen der KQ-Schätzung eines linearen Regressionsmodells.
- Wie lautet die Null- und Alternativhypothese für den White-Test und welche Schlussfolgerung würden Sie bei einem Signifikanzniveau von 5% ziehen?
 - Welche Annahme des Regressionsmodells aus Gleichung (1) wäre von Ihrem Ergebnis aus (i) betroffen bzw. müsste geändert werden?
Nennen Sie kurz die Konsequenzen dieser Annahmenänderung hinsichtlich der Erwartungstreue des KQ-Schätzers für β sowie der Gültigkeit der Inferenz (t - und F -Test).
- d) (3 Punkte) Beschreiben Sie, wie Sie testen würden, ob eine Beobachtung ein Ausreißer ist.
- e) (6 Punkte) Mittels eines linearen Regressionsmodells wie in Gleichung (1) wird versucht, den Absatz von Modegeschäften (*sales*) durch Charakteristika wie die Verkaufsfläche (*ssize*), die Zahl der Geschäftsinhaber (*nown*), die geleisteten Arbeitsstunden (*hoursw*) sowie die Anzahl der Voll- bzw. Teilzeitbeschäftigten (*nfull* bzw. *npart*) zu erklären. Hierzu liegt ein Querschnittsdatensatz vom Umfang 400 vor. Die Abbildungen 1 und 2 zeigen die Ergebnisse der KQ-Schätzung für zwei verschiedene Modellspezifikationen.
- Wählen Sie einer der beiden Spezifikationen. Vergleichen Sie hierzu die zwei Modelle auf Basis [1] des Akaike-Kriteriums, [2] des Schwarz-Kriteriums sowie [3] des adjustierten Bestimmtheitsmaßes (\bar{R}^2).
 - Das Akaike-Kriterium ist definiert als Summe von zwei Komponenten:

$$\text{AIC} := \ln\left(\frac{1}{N}\text{RSS}\right) + \frac{2K}{N} = \ln\left(\frac{1}{N}\sum_{i=1}^N e_i^2\right) + \frac{2K}{N}, \quad (2)$$

wobei e_i die KQ-Residuen $y_i - x_i'\hat{\beta}_{KQ}$ sind. Des Weiteren ist N die Anzahl der Beobachtungen und K die Anzahl der Regressoren.

Das Akaike-Kriterium charakterisiert die "Güte" eines Modells. Erläutern sie kurz, welche Funktion die beiden Komponenten des Kriteriums in diesem Zusammenhang haben. Gehen Sie dabei auch auf den sogenannten "Bias-Varianz-Trade-Off" ein.

Tabelle 1: White-Test

Statistik	P-Wert
8.151	0.043

Dependent Variable: SALES Method: Least Squares Date: 03/19/10 Time: 15:59 Sample: 1 400 Included observations: 400				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5133.590	321.6934	15.95802	0.0000
HOURSW	37.52842	2.837220	13.22718	0.0000
SSIZE	-22.14457	1.625067	-13.62686	0.0000
R-squared	0.365804	Mean dependent var		6334.751
Adjusted R-squared	0.362609	S.D. dependent var		3739.344
S.E. of regression	2985.371	Akaike info criterion		18.84831
Sum squared resid	3.54E+09	Schwarz criterion		18.87824
Log likelihood	-3766.662	F-statistic		114.4947
Durbin-Watson stat	1.957873	Prob(F-statistic)		0.000000

Abbildung 1: Schätzergebnisse - Spezifikation I

Dependent Variable: SALES Method: Least Squares Date: 03/19/10 Time: 16:54 Sample: 1 400 Included observations: 400				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3835.973	547.0482	7.012130	0.0000
NOWN	945.6822	255.0893	3.707259	0.0002
SSIZE	-15.79586	1.616579	-9.771167	0.0000
NFULL	1330.475	173.7459	7.657589	0.0000
NPART	586.0070	251.6179	2.328956	0.0204
R-squared	0.254794	Mean dependent var		6334.751
Adjusted R-squared	0.247247	S.D. dependent var		3739.344
S.E. of regression	3244.304	Akaike info criterion		19.01961
Sum squared resid	4.16E+09	Schwarz criterion		19.06950
Log likelihood	-3798.922	F-statistic		33.76361
Durbin-Watson stat	1.944006	Prob(F-statistic)		0.000000

Abbildung 2: Schätzergebnisse - Spezifikation II

Aufgabe 2 (14 Punkte)

- a) (4 Punkte) Unterstellen Sie folgendes lineares Regressionsmodell:

$$y = X\beta + \varepsilon \quad \text{bzw.} \quad y_i = x_i'\beta + \varepsilon_i; \quad i = 1, \dots, N, \quad (3)$$

wobei X eine stochastische $(N \times K)$ -Matrix und x_i ein stochastischer $(K \times 1)$ -Vektor ist.

Definieren Sie die Begriffe der [1] strikten Exogenität und [2] schwachen Exogenität der Regressoren.

Geben Sie ein Beispiel für eine Situation an, in der die Annahme der strikten Exogenität der Regressoren nicht erfüllt ist.

- b) Gegeben sei ein lineares Regressionsmodell für eine makroökonomische Konsumfunktion:

$$C_t = \beta_1 + \beta_2 Y_t + e_t, \quad e_t \sim (0, \sigma^2) \text{ i.i.d.} \quad (t = 1, \dots, T) \quad \text{bzw.} \quad C = Y\beta + e, \quad (4)$$

wobei C_t den realen Konsum und Y_t das reale Einkommen der Volkswirtschaft bezeichnet.

Gemäß der makroökonomischen Theorie sollte in einer geschlossenen Volkswirtschaft die Identität $Y_t = C_t + I_t$ erfüllt sein, wobei I_t die exogenen realen Investitionen bezeichnet.

- (i) (3 Punkte) Begründen Sie kurz, weshalb die Konsumfunktion nicht mittels der KQ-Methode geschätzt werden sollte.

Hinweis: Es genügt eine verbale Antwort. Eine formale Herleitung ist nicht erforderlich.

- (ii) (4 Punkte) Schlagen Sie einen Schätzer für den Parametervektor $\beta = (\beta_1, \beta_2)'$ vor und begründen Sie Ihre Auswahl kurz.

Geben Sie eine Formel für die Berechnung des von Ihnen vorgeschlagenen Schätzers in Matrixnotation an und erklären Sie die darin auftretenden Matrizen bzw. Vektoren.

- (iii) (3 Punkte) Welche Kriterien müssen Instrumentalvariablen idealerweise erfüllen? Erläutern Sie in diesem Zusammenhang kurz die Konzepte *Relevanz* von Instrumentalvariablen und *schwache Exogenität* von Instrumentalvariablen.

Aufgabe 3 (16 Punkte)

- a) (2 Punkte) Definieren Sie den Begriff der schwachen Stationarität eines stochastischen Prozesses.
- b) (6 Punkte) Mit dem Augmented-Dickey-Fuller-Test ("ADF-Test") wird überprüft, ob eine Zeitreihe Y_t integriert von der Ordnung 1 ist, d.h. eine $I(1)$ -Reihe ist. Die dem Test zugrundeliegende Regressionsgleichung lautet:

$$Y_t = \rho Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \dots + \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t, \quad (5)$$

äquivalent: $\Delta Y_t = \alpha Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \dots + \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t.$

wobei $\Delta Y_t := Y_t - Y_{t-1}$ und $\alpha := \rho - 1$.

Hinweis: Sie können sich bei der Beantwortung der Fragen für eine Formulierung der ADF-Regressionsgleichung (5) Ihrer Wahl entscheiden.

- (i) Formulieren Sie die Null- und Alternativhypothese des ADF-Tests bezüglich der Parameter der ADF-Regressionsgleichung (5).

Hinweis: Wählen Sie eine Formulierung von der Art $H_0: \theta = c$, wobei θ ein Parameter der Regressionsgleichung und c eine reelle Zahl ist.

- (ii) Wir unterscheiden beim ADF-Test drei Fälle: [1] ohne Konstante/ohne Trend, [2] mit Konstante/ohne Trend und [3] mit Konstante/mit Trend.

Erläutern Sie kurz, wie Sie in einer empirischen Untersuchung vorgehen würden, um zu entscheiden, ob Fall [1], [2] oder [3] des ADF-Tests verwendet werden sollte.

- (iii) Warum ist es überhaupt wichtig, Integrationsbeziehungen zu identifizieren und sich vor integrierten Zeitreihen zu „schützen“?

- c) (6 Punkte) Die Abbildungen 3 bzw. 4 zeigen die empirischen Autokorrelationen und partiellen Autokorrelationen für die zwei Zeitreihen X_t bzw. Y_t .

Schlagen Sie für beide Zeitreihen jeweils eine geeignete ARMA(p,q)-Spezifikation vor und begründen Sie Ihre Entscheidung.

Geben Sie für die von Ihnen gewählten ARMA(p,q)-Spezifikation jeweils die Modellgleichung an.

- d) (2 Punkte) Ist die KQ-Schätzung konsistent, wenn die Fehler autokorreliert sind? Geben Sie eine kurze Begründung.

Correlogram of X

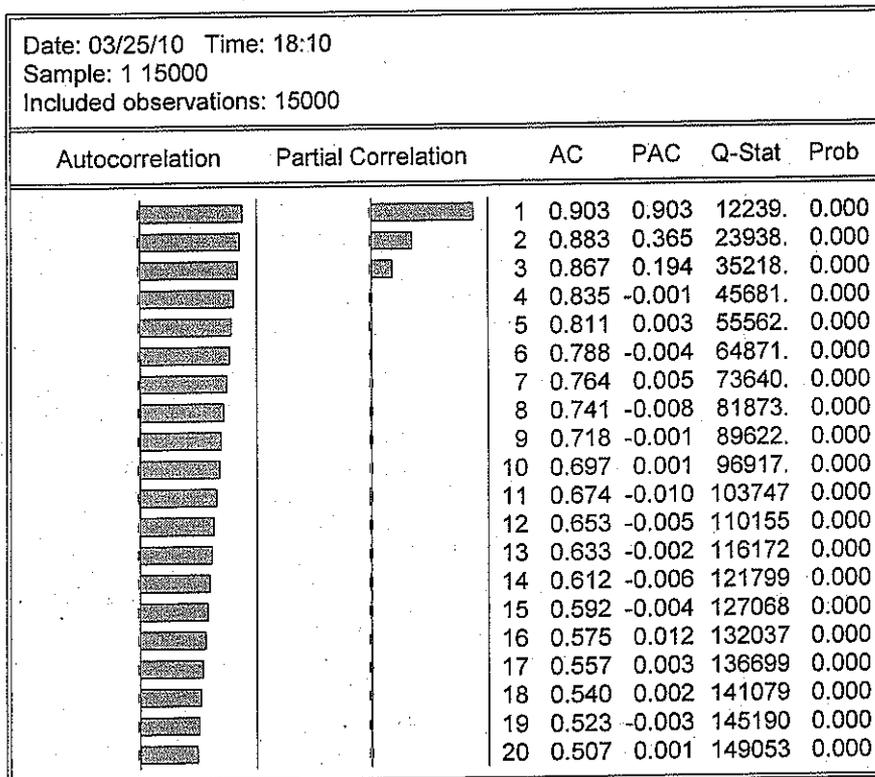


Abbildung 3: Empirische Autokorrelationen und partielle Autokorrelationen von X_t

Correlogram of Y

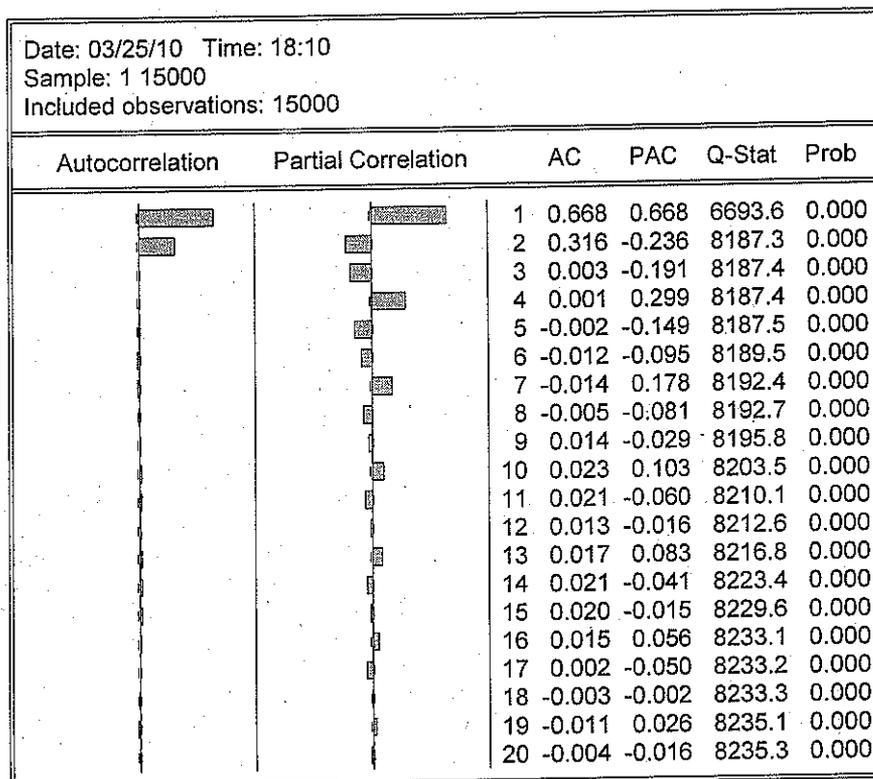


Abbildung 4: Empirische Autokorrelationen und partielle Autokorrelationen von Y_t