WS 09/10 II

Aufgabe 1 (15 Punkte)

Gegeben sei das folgende lineare Regressionsmodell:

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

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.

(i) Wie lautet die Null- und Alternativhypothese für den White-Test und welche Schlussfolgerung würden Sie bei einem Signifikanzniveau von 5% ziehen?

(ii) 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) (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. Teilzeitbeschätigten (`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.

(i) 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 ($\bar{R}^2$).

(ii) Das Akaike-Kriterium ist definiert als Summe von zwei Komponenten:

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

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:

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

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:

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

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) Erläutern Sie kurz die Grundidee eines Hausman-Tests. Gehen Sie dabei auch auf den Aufbau der Hausman-Teststatistik ein.

Aufgabe 3 (18 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:

$$Y_t = \rho\, Y_{t-1} + \zeta_1\, \Delta Y_{t-1} + \ldots + \zeta_{p-1}\, \Delta Y_{t-p+1} + \varepsilon_t, \tag{5}$$

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

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) Geben Sie an, welcher Typ von Teststatistik auf Basis der ADF-Regression (5) berechnet wird, um die Nullhypothese aus (i) zu überprüfen.

(iii) 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.

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.

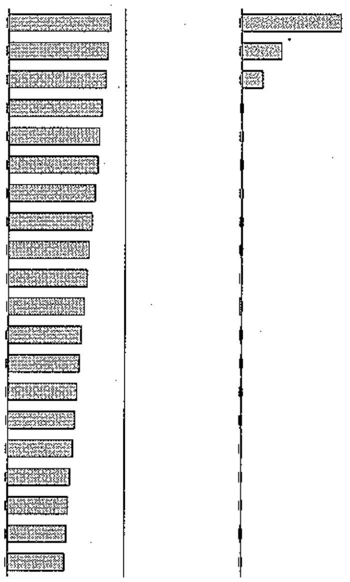
d) (4 Punkte) Gegeben sei ein Prozess $\{z_t\}$, der als schwach stationärer und kausaler AR(3)-Prozess mit einem Fehlerprozess $\{\varepsilon_t\} \sim \text{IWN}(0, \sigma^2)$ ("unabhängiges weißes Rauschen") modelliert werden kann. Wir nehmen an, dass Beobachtungen von z_t für T Perioden vorliegen ($t = 1, \ldots, T$) und der Wert z_{T+1} für Periode $T+1$ prognostiziert werden soll.

(i) Nehmen Sie an, dass die Koeffizienten des AR(3)-Modells für $\{z_t\}$ bekannt sind. Bestimmen Sie die auf dem bedingten Erwartungswert von z_{T+1} beruhende Ein-Schritt-Vorhersage $z_{T+1|T}$ (für z_{T+1}) zum Zeitpunkt T.

(ii) Die auf dem bedingten Erwartungswert von z_{T+1} beruhende Vorhersage $z_{T+1|T}$ aus (i) ist optimal ("bester Prädiktor") bezüglich eines bestimmten Kriteriums. Nennen Sie dieses Kriteriums und geben Sie dessen formalen Ausdruck an.

Correlogram of X

Date: 03/25/10 Time: 18:10
Sample: 1 15000
Included observations: 15000

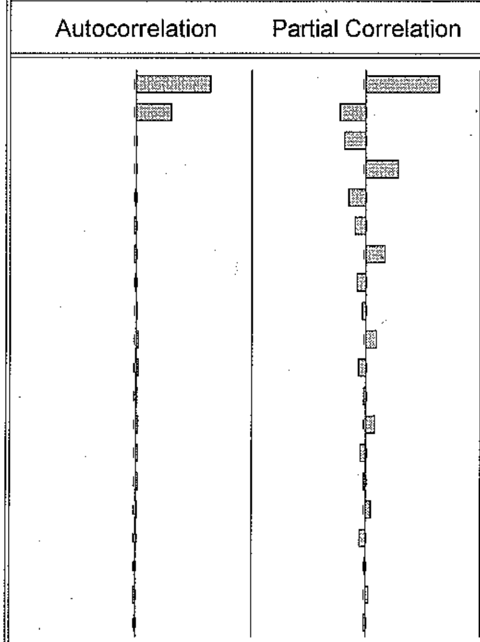


Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.903	0.903	12239.	0.000
		2	0.883	0.365	23938.	0.000
		3	0.867	0.194	35218.	0.000
		4	0.835	-0.001	45681.	0.000
		5	0.811	0.003	55562.	0.000
		6	0.788	-0.004	64871.	0.000
		7	0.764	0.005	73640.	0.000
		8	0.741	-0.008	81873.	0.000
		9	0.718	-0.001	89622.	0.000
		10	0.697	0.001	96917.	0.000
		11	0.674	-0.010	103747	0.000
		12	0.653	-0.005	110155	0.000
		13	0.633	-0.002	116172	0.000
		14	0.612	-0.006	121799	0.000
		15	0.592	-0.004	127068	0.000
		16	0.575	0.012	132037	0.000
		17	0.557	0.003	136699	0.000
		18	0.540	0.002	141079	0.000
		19	0.523	-0.003	145190	0.000
		20	0.507	0.001	149053	0.000

Abbildung 3: Empirische Autokorrelationen und partielle Autokorrelationen von X_t

Correlogram of Y

Date: 03/25/10 Time: 18:10
Sample: 1 15000
Included observations: 15000



	AC	PAC	Q-Stat	Prob
1	0.668	0.668	6693.6	0.000
2	0.316	-0.236	8187.3	0.000
3	0.003	-0.191	8187.4	0.000
4	0.001	0.299	8187.4	0.000
5	-0.002	-0.149	8187.5	0.000
6	-0.012	-0.095	8189.5	0.000
7	-0.014	0.178	8192.4	0.000
8	-0.005	-0.081	8192.7	0.000
9	0.014	-0.029	8195.8	0.000
10	0.023	0.103	8203.5	0.000
11	0.021	-0.060	8210.1	0.000
12	0.013	-0.016	8212.6	0.000
13	0.017	0.083	8216.8	0.000
14	0.021	-0.041	8223.4	0.000
15	0.020	-0.015	8229.6	0.000
16	0.015	0.056	8233.1	0.000
17	0.002	-0.050	8233.2	0.000
18	-0.003	-0.002	8233.3	0.000
19	-0.011	0.026	8235.1	0.000
20	-0.004	-0.016	8235.3	0.000

Abbildung 4: Empirische Autokorrelationen und partielle Autokorrelationen von Y_t