

WS 2010/11

Termin I

Aufgabe 1 (19 Punkte)

Gegeben sei das folgende lineare Regressionsmodell:

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

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

- a) (1 Punkt) Welche der zugrunde liegenden Modellannahmen ist im Fall einer exakten Multikollinearität verletzt?
- b) Eine KQ-Schätzung des Modells in Gleichung (1) liefert den Vektor der angepassten Werte $\hat{y} := X\hat{\beta}_{KQ}$ sowie den Residuenvektor $e := y - \hat{y}$, wobei $\hat{\beta}_{KQ}$ der KQ-Schätzer ist.

- (i) (3 Punkte) Abbildung 1 zeigt das Streudiagramm der angepassten Werte und standardisierten Residuen

$$\tilde{e}_i := \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}, \quad i = 1, \dots, N,$$

wobei $\hat{\sigma}^2 := \sum_{i=1}^N e_i^2 / (N - K)$ und $h_{ii} := x_i'(X'X)^{-1}x_i$ mit $x_i := (x_{i1}, \dots, x_{iK})'$.

Welche Schlüsse lässt das Streudiagramm hinsichtlich einer möglichen Heteroskedastizität der Fehlerterme des Regressionsmodells (1) zu? Begründen Sie Ihre Antwort.

- (ii) (1 Punkte) Nennen Sie einen Test, mit dem sich die Homoskedastizitätsannahme für die Fehlerterme auch überprüfen ließe.
- (iii) (8 Punkte) Nehmen Sie für diese Teilaufgabe an, die Regressormatrix X in Modell (1) sei stochastisch. Ferner gelte die generalisierte Fehlertermstruktur:

$$E[\varepsilon|X] = 0, \quad V[\varepsilon|X] = E[\varepsilon\varepsilon'|X] = \Psi = \sigma^2\Omega \neq \sigma^2I_N, \quad \varepsilon|X \sim N(0, \Psi), \quad (2)$$

wobei Ω eine symmetrische und positiv-definite Matrix sei.

[1] Bestimmen Sie den bedingten Erwartungswert $E[\hat{\beta}_{KQ}|X]$ sowie die bedingte Varianz $V[\hat{\beta}_{KQ}|X]$ des KQ-Schätzers $\hat{\beta}_{KQ}$.

[2] Nennen Sie die Folgen, welche Ihr Ergebnis aus [1] für die Verteilung des t- und F-Tests sowie bezüglich der Erwartungstreue und BLUE-Eigenschaft des KQ-Schätzers hat.

[3] Nennen Sie einen Schätzer für $V[\hat{\beta}_{KQ}|X]$, welcher unter der generalisierten Fehlertermstruktur (2) in jedem Fall konsistent ist.

- c) (6 Punkte) Nehmen Sie an, Gleichung (1) stelle das wahre Modell dar, während für die KQ-Schätzung die folgende Spezifikation verwendet wird:

$$y = X\beta + Z\gamma + u, \quad u \sim N(0, \sigma^2 I_N),$$

wobei Z eine deterministische $(N \times L)$ -Matrix sei mit $\text{Rang}(X:Z) = K + L$.

- (i) Bestimmen Sie den Erwartungswert $E[\hat{\beta}_{KQ}]$ des KQ-Schätzers $\hat{\beta}_{KQ} := (X'MX)^{-1}X'My$, wobei $M := I_N - Z(Z'Z)^{-1}Z'$.

Ist der KQ-Schätzer erwartungstreu?

- (ii) Bestimmen Sie die Varianz $V[\hat{\beta}_{KQ}]$ des KQ-Schätzers $\hat{\beta}_{KQ}$.

Nennen Sie die Folgen, welche Ihr Ergebnis bezüglich der BLUE-Eigenschaft des KQ-Schätzers hat.

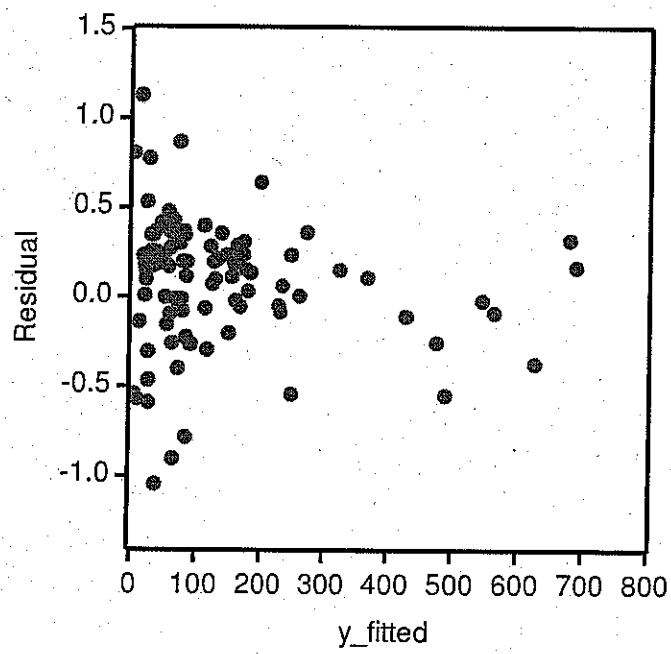


Abbildung 1: Streudiagramm der standardisierten Residuen ("Residual") und angepassten Werte ("y_fitted")

Aufgabe 2 (20 Punkte)

a) Gegeben sei folgendes Einweg-Fehlerkomponenten-Regressionsmodell:

$$y_{it} = \alpha + x'_{it} \beta + \varepsilon_{it}, \quad \varepsilon_{it} = \mu_i + e_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

wobei $x_{it} = (x_{1it}, \dots, x_{Kit})'$ (nicht-stochastisch) und $\beta = (\beta_1, \dots, \beta_K)'$ gilt. Des Weiteren seien alle μ_i und e_{it} unabhängige Zufallsgrößen mit $\mu_i \sim (0, \sigma_\mu^2)$ und $e_{it} \sim (0, \sigma_e^2)$. In Matrix-Notation lässt sich das Modell schreiben als

$$y = \alpha \mathbb{I}_{NT} + X\beta + \varepsilon = Z\theta + \varepsilon, \quad \varepsilon = G\mu + e,$$

wobei y , ε und e die NT -Vektoren aller Beobachtungen y_{it} und Fehler ε_{it} bzw. e_{it} sind, X die $NT \times K$ -Matrix der erklärenden Variablen mit den Zeilen x'_{it} ist sowie $\mu = (\mu_1, \dots, \mu_N)'$ und $G = I_N \otimes \mathbb{I}_T$ gilt. Darüber hinaus gilt $Z = [\mathbb{I}_{NT}; X]$ und $\theta = (\alpha \ \beta)'$.

(i) (4 Punkte) Welche Schätzung für β würden Sie im Falle bekannter Varianzkomponenten σ_μ^2 und σ_e^2 verwenden. Erklären Sie kurz, wie der Schätzer konzeptionell aufgebaut ist.

Hinweis: Die Angabe einer exakten Formel ist nicht erforderlich.

(ii) (6 Punkte) Zeigen Sie, dass eine "gepoolte" KQ-Schätzung für β erwartungstreu ist.

(iii) (4 Punkte) Unterstellen Sie nun, x_{it} bzw. X sei stochastisch. Erklären Sie, inwiefern zur erwartungstreuen (und konsistenten) Schätzung von β mit dem verallgemeinerten KQ-Ansatz (GLS) stärkere Annahmen benötigt werden als bei Verwendung des "Within"-Schätzers (Feste-Effekte-Schätzer).

b) Mit Hilfe von Paneldaten-Modellen soll untersucht werden, inwieweit die Variablen Berufserfahrung, Gewerkschaftsmitgliedschaft, Eheverhältnis und Anstellung im öffentlichen Sektor den Lohn eines Arbeitnehmers beeinflussen. Dazu liegt ein Paneldatensatz mit folgenden Variablen vor:

Variable	Beschreibung
wage	Logarithmierter Lohn
exper	Berufserfahrung
union	Dummy, =1 falls Person Mitglied einer Gewerkschaft ist
mar	Dummy, =1 falls Person verheiratet ist
pub	Dummy, =1 falls Person im öffentlichen Sektor angestellt ist

Unterstellt wird die Beziehung:

$$\text{wage}_{it} = \alpha + \beta_1 \text{exper}_{it} + \beta_2 \text{exper}_{it}^2 + \beta_3 \text{union}_{it} + \beta_4 \text{mar}_{it} + \beta_5 \text{pub}_{it} + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

(i) (2 Punkte) Abbildung 2 zeigt die Ergebnisse für die verallgemeinerte KQ-Schätzung (GLS). Unterstellen Sie das Fehlerkomponentenmodell:

$$\varepsilon_{it} = \mu_i + e_{it}, \quad \mu_i \sim (0, \sigma_\mu^2) \text{ i.i.d.}, \quad e_{it} \sim (0, \sigma_e^2) \text{ i.i.d.}$$

Geben Sie Schätzungen für σ_μ^2 und σ_e^2 auf Basis der Ergebnisse in Abbildung 2 an.

(ii) (4 Punkte) Die Tabellen 1 bzw. 2 zeigen die Ergebnisse für den Hausman- bzw. Breusch-Pagan-Test.

Geben Sie für beide Tests die Nullhypothese an und interpretieren Sie die Resultate hinsichtlich der Verwendung des verallgemeinerten KQ-Schätzers (GLS).

Hinweis: Legen Sie jeweils ein Signifikanzniveau von 5% zugrunde.

Dependent Variable: WAGE				
Method: Panel EGLS (Cross-section random effects)				
Date: 02/03/10 Time: 17:39				
Sample: 1980 1987				
Cross-sections included: 545				
Total panel (balanced) observations: 4360				
Swamy and Arora estimator of component variances				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.068730	0.030322	35.24653	0.0000
EXPER	0.117079	0.008254	14.18422	0.0000
EXPER^2	-0.004775	0.000589	-8.113367	0.0000
UNION	0.098799	0.017964	5.499727	0.0000
MAR	0.074532	0.016840	4.425810	0.0000
PUB	0.041135	0.036629	1.123022	0.2615
Effects Specification				
Cross-section random S.D. / Rho			0.351703	0.5006
Idiosyncratic random S.D. / Rho			0.351264	0.4994
Weighted Statistics				
R-squared	0.154117	Mean dependent var	0.549106	
Adjusted R-squared	0.153146	S.D. dependent var	0.384883	
S.E. of regression	0.354187	Sum squared resid	546.2035	
F-statistic	158.6570	Durbin-Watson stat	1.569466	
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.070681	Mean dependent var	1.649147	
Sum squared resid	1149.131	Durbin-Watson stat	0.745997	

Abbildung 2: Ergebnisse - Verallgemeinerte KQ-Schätzung (GLS)

Tabelle 1: Hausman-Test

Statistik	P-Wert
10.496	0.062

Tabelle 2: Breusch-Pagan-Test

Statistik	P-Wert
3217.140	0.000

Aufgabe 3 (20 Punkte)

- a) (2 Punkte) Welche Eigenschaften besitzt ein stochastischer Prozess, der als "Weißes Rauschen" ("White Noise") bezeichnet wird?

Die folgenden Fragen beziehen sich auf die Zeitreihe Z_t (mit $t = 1, \dots, 200$), welche in Abbildung 3 dargestellt ist.

- b) (6 Punkte) Überprüfen Sie, ob Z_t ein Einheitswurzelprozess, d.h. integriert von der Ordnung 1 ist ($Z_t \sim I(1)$). Gehen Sie dabei für die Analyse der Reihe Z_t bzw. ΔZ_t (erste Differenzen von Z_t) wie folgt vor:
- Entscheiden Sie sich unter Verwendung der graphischen Darstellungen der Reihen in den Abbildungen 3 bzw. 7 jeweils für einen geeigneten Fall des ADF-Tests und begründen Sie ihre Auswahl ([1] ohne Konstante und Trend, [2] mit Konstante, [3] mit Konstante und Trend).
 - Treffen Sie eine Testentscheidung auf Basis des ADF-Resultats für den in (i) gewählten Fall (Abbildungen 4, 5, 6, 8, 9 bzw. 10).

Wie lauten die Null- und Alternativhypothese des ADF-Tests?

- c) (2 Punkte) Abbildung 11 zeigt die empirischen Autokorrelationen und partiellen Autokorrelationen für die ersten Differenzen der Reihe Z_t .

Begründen Sie kurz, weshalb ein AR(3)-Prozess (und nicht ein anderer AR-Prozess) für die Modellierung herangezogen werden sollte.

- d) (4 Punkte) Abbildung 12 zeigt das Korrelogramm der Residuen nach Anpassung eines AR(3)-Modells für ΔZ_t .

- Welche Schlussfolgerungen können Sie aus dem Korrelogramm bzgl. der Anpassungsgüte des AR(3)-Modells ziehen? Geben Sie eine kurze Begründung an.
- Erklären Sie kurz auf intuitiver Basis, welche Überlegung der Analyse des Residuenkorrelogramms für ein geschätztes ARMA-Modell zugrunde liegt.

- e) (6 Punkte) Wir unterstellen ein AR(3)-Modell für ΔZ_t :

$$\Delta Z_t = c + \phi_1 \Delta Z_{t-1} + \phi_2 \Delta Z_{t-2} + \phi_3 \Delta Z_{t-3} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2),$$

wobei angenommen wird, dass die Lösungen z_1, z_2, z_3 von $(1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3) = 0$ außerhalb des Einheitskreises liegen, d.h. $|z_j| > 1$ für $j = 1, 2, 3$.

- Ist bei einer KQ-Schätzung dieses autoregressiven Modells die Annahme strikter Exogenität der Regressoren $\Delta Z_{t-1}, \Delta Z_{t-2}, \Delta Z_{t-3}$ erfüllt? Begründen Sie Ihre Antwort analytisch.
- Ist die KQ-Schätzung der AR-Parameter ϕ_1, ϕ_2, ϕ_3 konsistent? Geben Sie auch hier eine analytische Begründung an.

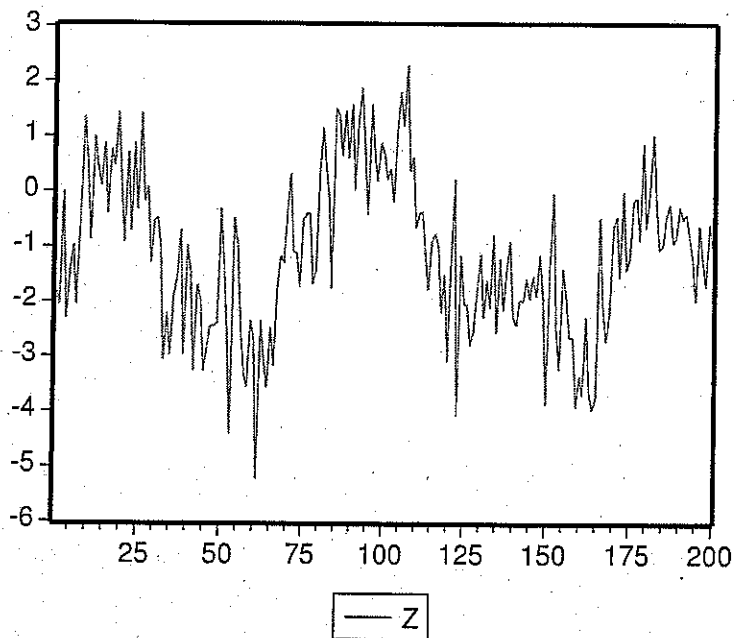


Abbildung 3: Z_t

Exogenous: Constant Lag Length: 3 (Automatic based on SIC, MAXLAG=14)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-2.345785	0.1588
Test critical values:				
	1% level		-3.463749	
	5% level		-2.876123	
	10% level		-2.574622	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Z)				
Method: Least Squares				
Date: 01/26/10 Time: 12:52				
Sample (adjusted): 5 200				
Included observations: 196 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.135578	0.057796	-2.345785	0.0200
D(Z(-1))	-0.566864	0.081029	-6.995831	0.0000
D(Z(-2))	-0.448014	0.080291	-5.579898	0.0000
D(Z(-3))	-0.220477	0.069854	-3.156251	0.0019
C	-0.141944	0.094773	-1.497731	0.1359
R-squared	0.353935	Mean dependent var		0.006661
Adjusted R-squared	0.340405	S.D. dependent var		1.217694
S.E. of regression	0.988956	Akaike info criterion		2.840845
Sum squared resid	186.8045	Schwarz criterion		2.924471
Log likelihood	-273.4028	F-statistic		26.15897
Durbin-Watson stat	2.031433	Prob(F-statistic)		0.000000

Abbildung 4: ADF-Test für Z_t (mit Konstante)

Exogenous: None				
Lag Length: 3 (Automatic based on SIC, MAXLAG=14)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.801765	0.0682
Test critical values:	1% level		-2.576814	
	5% level		-1.942456	
	10% level		-1.615622	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Z)				
Method: Least Squares				
Date: 01/26/10 Time: 12:56				
Sample (adjusted): 5 200				
Included observations: 196 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.077869	0.043218	-1.801765	0.0732
D(Z(-1))	-0.609716	0.076054	-8.016829	0.0000
D(Z(-2))	-0.476918	0.078189	-6.099573	0.0000
D(Z(-3))	-0.235363	0.069367	-3.393026	0.0008
R-squared	0.346347	Mean dependent var		0.006661
Adjusted R-squared	0.336134	S.D. dependent var		1.217694
S.E. of regression	0.992153	Akaike info criterion		2.842317
Sum squared resid	188.9984	Schwarz criterion		2.909218
Log likelihood	-274.5471	Durbin-Watson stat		2.039881

Abbildung 5: ADF-Test für Z_t (ohne Konstante)

Exogenous: Constant, Linear Trend				
Lag Length: 3 (Automatic based on SIC, MAXLAG=14)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-2.382869	0.3873
Test critical values:	1% level		-4.005562	
	5% level		-3.432917	
	10% level		-3.140265	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Z)				
Method: Least Squares				
Date: 01/26/10 Time: 12:57				
Sample (adjusted): 5 200				
Included observations: 196 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z(-1)	-0.139320	0.058467	-2.382869	0.0182
D(Z(-1))	-0.563987	0.081429	-6.926142	0.0000
D(Z(-2))	-0.446567	0.080515	-5.546364	0.0000
D(Z(-3))	-0.219860	0.070010	-3.140400	0.0020
C	-0.086184	0.152605	-0.564757	0.5729
@TREND(1)	-0.000590	0.001263	-0.466790	0.6412
R-squared	0.354675	Mean dependent var		0.006661
Adjusted R-squared	0.337693	S.D. dependent var		1.217694
S.E. of regression	0.990987	Akaike info criterion		2.849903
Sum squared resid	186.5905	Schwarz criterion		2.950254
Log likelihood	-273.2905	F-statistic		20.88506
Durbin-Watson stat	2.031931	Prob(F-statistic)		0.000000

Abbildung 6: ADF-Test für Z_t (mit Konstante und Trend)

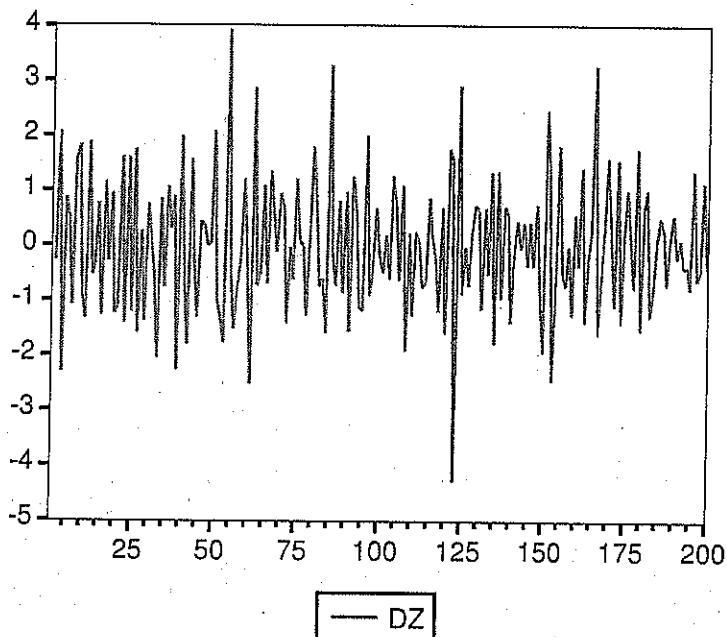


Abbildung 7: Erste Differenzen von Z_t (ΔZ_t)

Exogenous: Constant Lag Length: 2 (Automatic based on SIC, MAXLAG=14)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-14.04905	0.0000
Test critical values:				
	1% level		-3.463749	
	5% level		-2.876123	
	10% level		-2.574622	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DZ)				
Method: Least Squares				
Date: 01/26/10 Time: 12:58				
Sample (adjusted): 5 200				
Included observations: 196 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DZ(-1)	-2.438277	0.173555	-14.04905	0.0000
D(DZ(-1))	0.770571	0.126880	6.073214	0.0000
D(DZ(-2))	0.254988	0.069083	3.691013	0.0003
C	0.006266	0.071464	0.087674	0.9302
R-squared	0.767808	Mean dependent var		0.009652
Adjusted R-squared	0.764180	S.D. dependent var		2.060253
S.E. of regression	1.000485	Akaike info criterion		2.859044
Sum squared resid	192.1863	Schwarz criterion		2.925944
Log likelihood	-276.1863	F-statistic		211.6343
Durbin-Watson stat	2.049244	Prob(F-statistic)		0.000000

Abbildung 8: ADF-Test für erste Differenzen von Z_t (mit Konstante)

Exogenous: None Lag Length: 2 (Automatic based on SIC, MAXLAG=14)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-14.08519	0.0000
Test critical values:				
	1% level		-2.576814	
	5% level		-1.942456	
	10% level		-1.615622	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DZ)				
Method: Least Squares				
Date: 01/26/10 Time: 12:59				
Sample (adjusted): 5 200				
Included observations: 196 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DZ(-1)	-2.438255	0.173108	-14.08519	0.0000
D(DZ(-1))	0.770544	0.126553	6.088692	0.0000
D(DZ(-2))	0.254972	0.068905	3.700324	0.0003
R-squared	0.767799	Mean dependent var		0.009652
Adjusted R-squared	0.765393	S.D. dependent var		2.060253
S.E. of regression	0.997910	Akaike info criterion		2.848880
Sum squared resid	192.1940	Schwarz criterion		2.899055
Log likelihood	-276.1902	Durbin-Watson stat		2.049147

Abbildung 9: ADF-Test für erste Differenzen von Z_t (ohne Konstante)

Exogenous: Constant, Linear Trend Lag Length: 2 (Automatic based on SIC, MAXLAG=14)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-14.01380	0.0000
Test critical values:				
	1% level		-4.005562	
	5% level		-3.432917	
	10% level		-3.140265	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DZ)				
Method: Least Squares				
Date: 01/26/10 Time: 13:00				
Sample (adjusted): 5 200				
Included observations: 196 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DZ(-1)	-2.438475	0.174005	-14.01380	0.0000
D(DZ(-1))	0.770797	0.127216	6.058974	0.0000
D(DZ(-2))	0.255088	0.069264	3.682830	0.0003
C	0.024229	0.147170	0.164632	0.8694
@TREND(1)	-0.000177	0.001266	-0.139736	0.8890
R-squared	0.767832	Mean dependent var		0.009652
Adjusted R-squared	0.762970	S.D. dependent var		2.060253
S.E. of regression	1.003049	Akaike info criterion		2.869146
Sum squared resid	192.1667	Schwarz criterion		2.952771
Log likelihood	-276.1763	F-statistic		157.9201
Durbin-Watson stat	2.049528	Prob(F-statistic)		0.000000

Abbildung 10: ADF-Test für erste Differenzen von Z_t (mit Konstante und Trend)

Correlogram of DZ

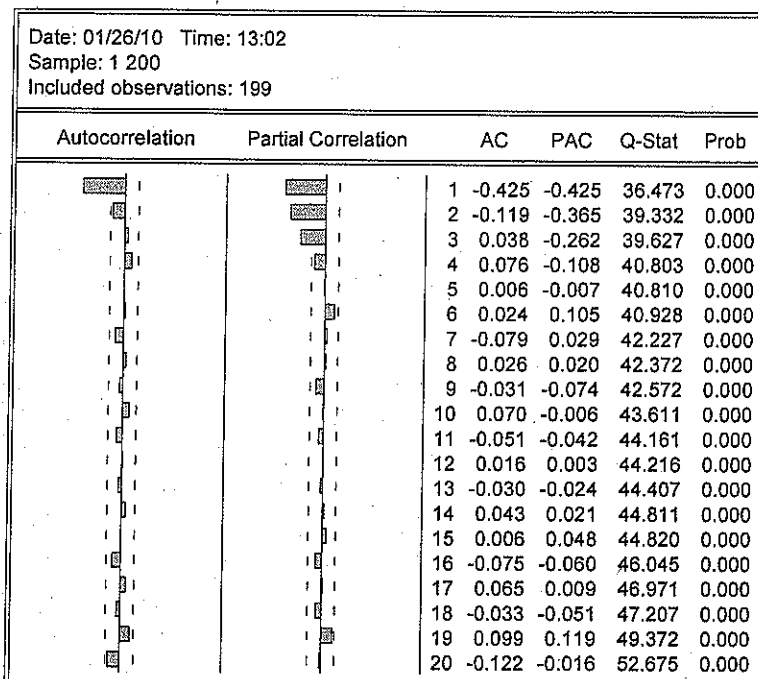


Abbildung 11: Empirische Autokorrelationen und partielle Autokorrelationen der ersten Differenzen von Z_t

Correlogram of Residuals

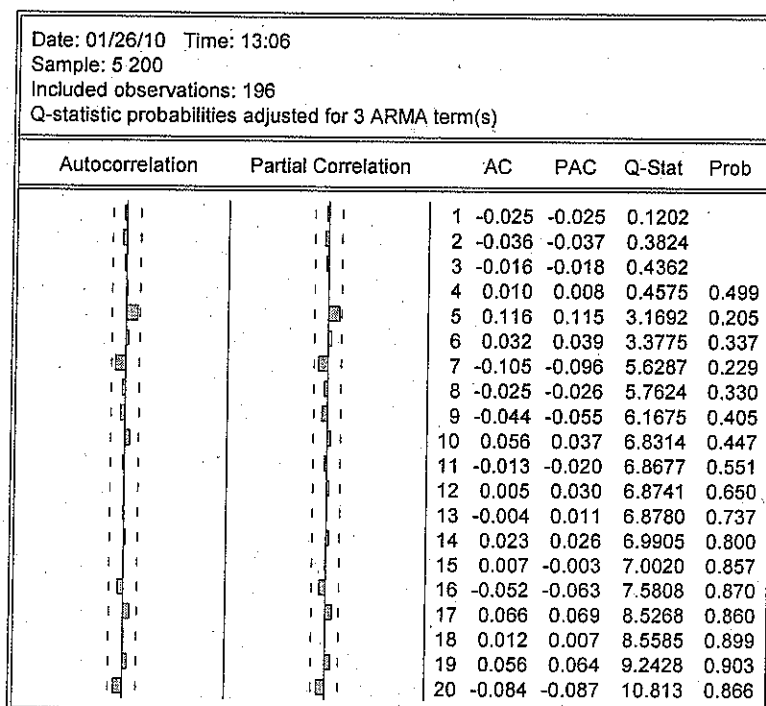


Abbildung 12: Empirische Autokorrelationen und partielle Autokorrelationen der Residuen