Prüfung im Fach Ökonometrie im WS 2010/11 – Aufgabenteil

Name, Vorname	
Matrikelnr.	
Studiengang	
E-Mail-Adresse	
Unterschrift	

Vorbemerkungen:

Anzahl der Aufgaben:
- Die Klausur besteht aus 4 Aufgaben.

Bewertung:
- Die Prüfungsdauer beträgt für alle Studierenden 90 Minuten, es können maximal 90 Punkte erworben werden.
- Die Punktzahl für jede Aufgabe ist in Klammern angegeben.

Erlaubte Hilfsmittel:
- Tabellen der statistischen Verteilungen und Formelsammlung (sind der Klausur beigefügt)
- Taschenrechner
- Fremdwörterbuch

Wichtige Hinweise:
- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den exakten Wert der gesuchten Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
- Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Annahme oder Angabe fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.
- Aufgabe 4 ist im Aufgabenteil zu beantworten, die restlichen Aufgaben im Lösungsteil. Verwenden Sie hierbei für jede Aufgabe ein neues Blatt.

Aufgabe 1 (17.5 Punkte)

Der Marktwert von Gebrauchtwagen wurde mit Daten von $N = 804$ PKWs geschätzt. Tabelle 1 enthält die deskriptiven Statistiken und eine Beschreibung der Variablen. Die Schätzergebnisse des Modells sind in Tabelle 2 angegeben.

Tabelle 1: Deskriptive Statistiken

Variable	**Mittelwert**	**Std. Abw.**	**Min.**	**Max.**	**Beschreibung**
price	21343	9885	8639	70755	Marktwert in USD
ln_mileage	9.75	0.65	5.58	10.83	Fahrleistung in Meilen (logarithmiert)
cylinder6	0.39	0.49	0	1	=1, falls 6-Zylinder-Motor; 0 sonst (Referenz: 4-Zylinder-Motor)
cylinder8	0.12	0.33	0	1	=1, falls 8-Zylinder-Motor; 0 sonst (Referenz: 4-Zylinder-Motor)
doors4	0.76	0.43	0	1	=1, falls Viertürer; 0 sonst
cruise	0.75	0.43	0	1	=1, falls Fahrzeug Tempomat hat; 0 sonst
sound	0.68	0.47	0	1	=1, falls Sonderausstattung mit Sound-System; 0 sonst

Tabelle 2: Regressionsergebnisse

```
      Source |       SS       df       MS              Number of obs =     804
-------------+------------------------------           F(  6,   797) =  168.03
       Model |  4.3820e+10     6  7.3033e+09           Prob > F      =  0.0000
    Residual |  3.4641e+10   797  43464643.6           R-squared     =  0.5585
-------------+------------------------------           Adj R-squared =  0.5552
       Total |  7.8461e+10   803    97710315           Root MSE      =  6592.8

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ln_mileage |  -2069.477   360.0708    -5.75   0.000    -2776.276   -1362.678
   cylinder6 |   514.7712   533.4205     0.97   0.335    -532.3038    1561.846
   cylinder8 |   18058.42   780.0165    23.15   0.000     16527.29    19589.54
      doors4 |  -1043.097   564.6848    -1.85   0.065    -2151.542    65.34796
      cruise |   6701.267   581.1873    11.53   0.000     5560.428    7842.105
       sound |  -570.5767   505.7804    -1.13   0.260    -1563.396    422.2424
       _cons |   35225.15   3603.068     9.78   0.000     28152.53    42297.77
------------------------------------------------------------------------------
```

1.1 Interpretieren Sie statistisch und inhaltlich den Zusammenhang zwischen der Fahrleistung (`ln_mileage`) und dem Preis eines Fahrzeugs. (1.5 Punkte)

1.2 Berechnen und interpretieren Sie das 90%-Konfidenzintervall für den Koeffizienten von `ln_mileage`. (3 Punkte)

1.3 Der folgende Stata-Output enthält die Ergebnisse eines Breusch-Pagan-Tests auf Heteroskedastizität. Führen Sie den Test am Signifikanzniveau $\alpha = 0.01$ durch. Geben Sie die Hypothesen, Teststatistik (inkl. Hilfsregression), kritischen Wert sowie die Testentscheidung an. (5 Punkte)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of price

         chi2(6)      =   168.30
         Prob > chi2  =   0.0000
```

1.4 Nennen Sie drei Konsequenzen von Heteroskedastizität für eine KQ-Schätzung und die anschließende Inferenz. (3 Punkte)

1.5 Erläutern Sie ausgehend von der Formel $Var(\mathbf{b}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Diag\{\sigma_i^2\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ kurz die Idee von White (1980) zur Berechnung robuster Standardfehler und legen Sie dar, wie diese durchgeführt wird. (3 Punkte)

1.6 In welcher Situation ist die Berechnung robuster Standardfehler als Lösung für das Problem heteroskedastischer Störterme dem (F)GLS-Schätzer vorzuziehen? (2 Punkte)

Aufgabe 2 (27.5 Punkte)

Auf Basis von Zeitreihendaten wurde die Nachfragefunktion von Konsumenten auf einem Wochenmarkt geschätzt. Die Modellspezifikation lautet:

$$\ln Q_t = \beta_0 + \beta_1 \ln p_t + \beta_2 Mon_t + \beta_3 Tues_t + \beta_4 Wed_t + \beta_5 Thur_t + \nu_t \ .$$

Q_t bezeichnet die nachgefragte Menge (in Kilogramm), p_t ist der Preis zum Zeitpunkt t. *Mon* bis *Thur* sind Dummy-Variablen für die Wochentage Montag bis Donnerstag. Die Referenzkategorie ist Freitag. ν_t bezeichnet den Störterm. Die Ergebnisse einer KQ-Schätzung sind in Tabelle 3 wiedergegeben.

Tabelle 3: Schätzergebnisse für die Nachfragefunktion

```
      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  5,    91) =    3.75
       Model |  13.0514133     5  2.61028266           Prob > F      =  0.0039
    Residual |  63.2947118    91  .695546284           R-squared     =  0.1710
-------------+------------------------------           Adj R-squared =  0.1254
       Total |  76.3461251    96  .795272137           Root MSE      =  .83399

------------------------------------------------------------------------------
        ln_Q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_p |  -.5667372   .2017012    -2.81   0.006    -.9673918   -.1660826
         mon |  -.5716882   .2711597    -2.11   0.038    -1.110314   -.0330628
        tues |  -.8045457   .2672322    -3.01   0.003     -1.33537   -.2737218
         wed |  -.4804791   .2637327    -1.82   0.072    -1.004352    .0433936
       thurs |  -.1426947   .2640372    -0.54   0.590    -.6671722    .3817828
       _cons |   7.769402   .1951757    39.81   0.000      7.38171    8.157095
------------------------------------------------------------------------------
```

2.1 Interpretieren Sie den Schätzer für den Koeffizienten β_1 statistisch und inhaltlich. (1.5 Punkte)

2.2 Erweitern Sie die obige Modellspezifikation, um zu testen, ob sich der Zusammenhang zwischen Nachfrage und Preis am Donnerstag von dem am Freitag unterscheidet. Spezifizieren Sie eine zur Überprüfung der Hypothese geeignete Schätzgleichung und geben Sie die Null- und Alternativhypothese sowie die Teststatistik an. (6 Punkte)

2.3 Erläutern Sie, warum die folgende Interpretation des Koeffizienten β_3 aus Tabelle 3 unpräzise ist: „Die Nachfrage ist im Mittel c. p. dienstags um 80.45% Prozent geringer als freitags." Geben Sie die exakte Interpretation an. (2 Punkte)

2.4 Gehen Sie von folgendem Regressionsmodell aus: $\ln Q_t = \beta_0 + \beta_1 \ln p_t + \varepsilon_t$. Führen Sie einen Breusch-Godfrey-Test auf Autokorrelation 1. Ordnung am Signifikanznivea $\alpha = 0.1$ durch. Geben Sie Hypothesen, Teststatistik (inkl. Hilfsregression), kritischen Wert und Testentscheidung an. *Hinweis*: Zur Durchführung des Test können Sie die Ergebnisse in Tabelle 4 verwenden. (5 Punkte)

Tabelle 4: Stata-Output für den Befehl -estout bgodfrey-

```
. estat bgodfrey, nomiss0 lags(1)

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          4.189               1                   0.0407
---------------------------------------------------------------------------
                        H0: no serial correlation
```

2.5 Die Nachfragefunktion $\ln Q_t = \beta_0 + \beta_1 \ln p_t + \varepsilon_t$ kann mit einem GLS-Schätzer geschätzt werden, wenn der Störterm ε_t einem autoregressiven Prozess der Form $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$ folgt und für den Störterm ν_t die Gauß-Markov-Annahmen erfüllt sind.

2.5.1 Geben Sie die GLS-Transformation des Modells für die Beobachtungen $t \geqslant 2$ an. Wie wird dieses Schätzverfahren bezeichnet? (2.5 Punkte)

2.5.2 Geben Sie die GLS-Transformation des Modells für die erste Beobachtung $t = 1$ an. Wie wird das Schätzverfahren bezeichnet, das die transformierten Beobachtungen aller Zeitpunkte $t = 1, 2, \ldots, T$ nutzt? (2.5 Punkte)

2.6 Stellen Sie die Varianz-Kovarianz-Matrix für einen *moving average* Prozess für den Fall $T = 3$ dar. Berechnen Sie dazu $Var(\varepsilon_t), Cov(\varepsilon_t, \varepsilon_{t-1})$ sowie $Cov(\varepsilon_t, \varepsilon_{t-2})$. Gehen Sie dabei von folgenden Annahmen aus: $\varepsilon_t = \nu_t + \nu_{t-1}$ und $Var(\nu_t) = \sigma_\nu^2$. Ferner sind die Gauß-Markov-Annahmen für ν_t erfüllt. (8 Punkte)

Aufgabe 3 (15 Punkte)

Die Wahrscheinlichkeit, eine Schiffskatastrophe zu überleben, wurde anhand von Daten über den Untergang der Titanic mit einem binären Modell analysiert (siehe Tabelle 6). Tabelle 5 beschreibt die verwendeten Variablen.

Tabelle 5: Deskriptive Statistiken

Variable	**Mittelwert**	**Std. Abw.**	**Min.**	**Max.**	**Beschreibung**
survived	0.32	0.47	0	1	=1, falls Person überlebt hat; 0 sonst.
child	0.05	0.22	0	1	=1, falls Person ein Kind ist; 0 sonst.
male	0.79	0.41	0	1	=1, falls Person ein erwachsener Mann ist; 0 sonst.
first_class	0.15	0.35	0	1	=1, falls Passagier der 1. Klasse; 0 sonst (Referenz: Besatzung)
second_class	0.13	0.34	0	1	=1, falls Passagier der 2. Klasse; 0 sonst (Referenz: Besatzung)
third_class	0.32	0.47	0	1	=1, falls Passagier der 3. Klasse; 0 sonst (Referenz: Besatzung)

Tabelle 6: Regressionsergebnisse

```
Logistic regression                               Number of obs   =       2201
                                                  LR chi2(5)      =     559.40
                                                  Prob > chi2     =     0.0000
Log likelihood = -1105.0306                       Pseudo R2       =     0.2020

------------------------------------------------------------------------------
    survived |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       child |   1.061542   .2440257     4.35   0.000     .5832608    1.539824
        male |   -2.42006   .1404101   -17.24   0.000    -2.695259   -2.144862
 first_class |   .8576762   .1573389     5.45   0.000     .5492976    1.166055
second_class |  -.1604188   .1737865    -0.92   0.356    -.5010342    .1801966
 third_class |  -.9200861   .1485865    -6.19   0.000     -1.21131   -.6288619
       _cons |   1.186161   .1585673     7.48   0.000     .8753751    1.496947
------------------------------------------------------------------------------
```

3.1 Erläutern Sie die Intuition des Maximum-Likelihood-Verfahrens. Um welchen konkreten Schätzer handelt es sich in Tabelle 6? (3 Punkte)

3.2 Berechnen Sie die Überlebenswahrscheinlichkeiten von erwachsenen Männern und von erwachsenen Frauen und bestimmen Sie deren Differenz. Unterstellen Sie Stichprobenmittelwerte für die Schiffsklassen. *Hinweis*: $P(y_i = 1|\mathbf{x}_i) = F(\mathbf{x}_i'\boldsymbol{\beta}) = \frac{exp\{\mathbf{x}_i'\boldsymbol{\beta}\}}{1+exp\{\mathbf{x}_i'\boldsymbol{\beta}\}} = \frac{1}{1+exp\{-\mathbf{x}_i'\boldsymbol{\beta}\}}$. (7 Punkte)

3.3 Überprüfen Sie die Hypothese, dass die Überlebenswahrscheinlichkeit für Besatzungsmitglieder und Passagiere in der 1., 2. und 3. Klasse ceteris paribus gleich ist. Führen Sie dazu einen Likelihood-Ratio Test am 1%-Signifikanzniveau durch. Geben Sie Hypothesen, Teststatistik, Entscheidungsregel (kritischen Wert) und Testentscheidung an. *Hinweis*: Der Log-Likelihoodwert des restringierten Modells ist $\log L(\tilde{\theta}) = -1164.5475$. (5 Punkte)

Aufgabe 4: Wahr-Falsch Fragen (30 Punkte)

Wahr oder falsch? Tragen Sie für jede der folgenden Aussagen ein „w“ für wahr oder ein „f“ für falsch ein. Für jede richtige Antwort gibt es 0,75 Punkte, für jede falsche Antwort werden 0,75 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

Der ML-Schätzer maximiert die Wahrscheinlichkeit der quadrierten Residuen.	
Die Inverse einer Matrix hat die gleiche Dimension wie die Ausgangsmatrix.	
Die t-Statistik kann nicht negativ werden.	
Heteroskedastie und Autokorrelation können auch gleichzeitig vorliegen.	
Die F-Verteilung ist eine symmetrische Verteilung.	
Bei perfekter Multikollinearität von $\mathbf{X}$ hat die Kreuzproduktmatrix $(\mathbf{X}'\mathbf{X})$ keinen vollen Rang.	
Bei einer Regression ohne Konstante β_0 verläuft die Regressionsgerade durch den Koordinatenursprung.	
Beim KQ-Schätzer ist der Vektor der Residuen immer orthogonal zum $\mathbf{x}$-Vektor.	
Bei Vorliegen von Autokorrelation ist die Varianz-Kovarianz-Matrix des Störterms keine diagonale Matrix.	
Jeder Parameterschätzer ist effizient, der gegen den wahren Wert des Parameters konvergiert.	
Da die log-Likelihoodfunktion global konvex ist, konvergieren die Schätzungen schnell zum Maximum.	
Die Annahme $E(\varepsilon_t\|\mathbf{x}_t) = 0$ lässt zu, dass $Cov(\varepsilon_t, \varepsilon_{t-1}) > 0$ ist.	
Nicht-perfekte Multikollinearität ist unproblematisch, wenn das Ziel der Regression ein Punktschätzer der Steigungsparameter ist.	
Solange bei heteroskedastischen Störtermen keine Autokorrelation vorliegt, sind KQ Schätzer effizient.	
Um die Konsistenz des Kleinstquadrateschätzers zu beweisen, braucht man stärkere Annahmen als zum Nachweis seiner Unverzerrtheit.	
Wenn der p-Wert größer ist als das Signifikanzniveau eines Tests, wird die Nullhypothese verworfen.	
Durch das Hinzufügen weiterer erklärender Variablen kann der angepasste R2 Wert nicht sinken.	
Unter den Annahmen A1-A4 sind die auf Basis linearer Modelle geschätzten Koeffizienten als kausale Effekte zu interpretieren.	
Das Cramer-Rao-lower-bound stellt die Untergrenze der Durbin-Watson-Teststatistik dar.	
Das Signifikanzniveau beim einseitigen t-Test variiert mit der Beobachtungszahl.	
Bei einem einseitigen t-Test gegen die Alternative H1: $\beta < 0$, befindet sich der Ablehnungsbereich auf der linken Seite der Verteilung.	
Der Durbin-Watson-Test verallgemeinert den Breusch-Godfrey-Test.	
Das Auslassen einer relevanten Variable z führt zu verzerrten Parameterschätzern für den Koeffizienten von x, wenn $Cov(x,z) > 0$.	
Für die Vorhersage von y_i spielt es keine Rolle, ob die abhängige Variable logarithmiert ist oder nicht.	
Mit einem in den Parametern linearen Modell lassen sich auch Elastizitäten berechnen.	
Beim Test einer linearen Restriktion kommen t- und F-Test zum gleichen Ergebnis.	
Der RESET-Test dient dazu, die unterstellte funktionale Form der Regressionsgleichung zu überprüfen.	
Ein Chow-Test für 2 Gruppen führt zum gleichen Ergebnis wie ein F-Test der Interaktionsterme in einem vollständig interagierten Modell.	

Durch Umskalieren einer unabhängigen Variable ändert sich der Wert der Konstanten.	
Der iterative Cochrane-Orcutt Schätzer wird bei Vorliegen von Heteroskedastizität verwendet.	
Das R^2 kann durch die Aufnahme zusätzlicher Variablen nicht sinken.	
Je größer der Typ II Fehler eines Tests, umso größer muss der Typ I Fehler sein.	
Bedingung für einen asymptotisch normalverteilten KQ-Schätzer sind normalverteilte Störterme.	
Identisch und unabhängig verteilte Störterme können heteroskedastisch sein.	
Die Teststatistik des Wald-Tests ist χ^2-verteilt.	
Ein R^2 von 0.6 bei einem GLS-Schätzer besagt, dass 60% der Variation in y durch das Modell erklärt werden.	
Homoskedastische Störterme haben eine Varianz von 0.	
Die Wald-Teststatistik, die Likelihood Ratio-Teststatistik und die LM Teststatistik sind asymptotisch χ^2 verteilt.	
Binäre abhängige Variablen sind im Interval $[0,1]$ normalverteilt.	
Negative Autokorrelation kommt typischerweise in Querschnittsdatensätzen vor.	