Musterklausur

16. Juli 2014

Name:

Matrikelnummer:

Die Klausur dauert 90 Minuten.

Die Klausur besteht aus vier Aufgaben.

Es können maximal 90 Punkte erreicht werden.

Bitte beantworten Sie alle Fragen kurz und präzise. Sie können Stichpunkte verwenden.

Es sind keine Hilfsmittel zugelassen.

Viel Erfolg!

Aufgabe 1

Geben Sie bei jeder der folgenden Aussagen an, ob sie wahr oder falsch ist. Für jede richtige Antwort erhalten Sie einen Punkt, für jede falsche wird ein Punkt abgezogen. Sie erhalten im schlechtesten Fall eine Gesamtpunktzahl von null für Aufgabe 1.

Aussage	Wahr	Falsch
Das Gesetz der großen Zahlen besagt, dass der Stichprobenmittelwert standardnormalverteilt ist, wenn n gegen unendlich geht.		
$\mathbb{E}(u \mid x)$ impliziert, dass $\text{Cov}(u, x) = 0$.		
Im Modell $y = \beta_0 + \beta_1 \ln x + u$ entspricht $\beta_1 \cdot 100$ der prozentualen Änderung von $\mathbb{E}(y \mid x)$, wenn x um eine Einheit ansteigt.		
Wenn sich eine erklärende Variable als exakte Linearkombination von zwei anderen erklärenden Variablen darstellen lässt, darf sie nicht im Modell berücksichtigt werden.		
Ein R^2 von 0.12 bedeutet, dass das Modell statistisch insignifikant ist.		
Wenn Z standardnormalverteilt ist und X einer χ^2-Verteilung mit m Freiheitsgraden folgt, dann ist $T \equiv Z/X$ t-verteilt mit m Freiheitsgraden.		
Die bedingte Varianz des KQ-Schätzers für β_j, $j = 1 \ldots, k$, ist umso größer, je größer die Gesamtvariation von x_j ist.		
Die bedingte Varianz des KQ-Schätzers für β_j, $j = 1 \ldots, k$, ist umso größer, je stärker x_j mit den anderen Regressoren korreliert.		
Das Gauss-Markov-Theorem besagt, dass die KQ-Methode unverzerrte und konsistente Schätzer in der Klasse der linearen Schätzer liefert.		
Der p-Wert gibt das kleinste Signifikanzniveau an, zu dem man die Nullhypothese annehmen kann.		
Wenn x_2 in dem Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ nicht beobachtbar ist, kann man β_1 konsistent schätzen mit einer Regression von y auf x_1 allein, sofern $\text{Cov}(x_1, x_2) = 0$ ist.		
Die KQ-Schätzer $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ für das Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ sind im Allgemeinen inkonsistent, wenn x_2 mit dem Störterm korreliert.		
Wenn man in einer Regression eine Männerdummy und Interaktionsterme zwischen der Männerdummy und allen anderen Regressoren einfügt, ist das äquivalent dazu getrennte Regressionen für Männer und Frauen durchzuführen.		
Heteroskedastie in dem Modell $y = \beta_0 + \beta_1 x + u$ kann darauf hindeuten, dass x nicht nur linear, sondern auch quadratisch eingehen sollte.		
Autokorrelation der Störterme hat zur Folge, dass die KQ-Schätzer verzerrt und inkonsistent sind.		

[15 Punkte]

Aufgabe 2

Gegeben sei das folgende Regressionsmodell, für das die Gauss-Markov-Annahmen (MLR.1-MLR.5) erfüllt sind:

$$\mathbf{y} \;=\; \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

wobei

$$\mathbf{y} \equiv \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} \equiv \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{pmatrix}, \quad \boldsymbol{\beta} \equiv \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \text{und} \;\; \mathbf{u} \equiv \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

a) Zeigen Sie, dass der KQ-Schätzer die Lösung des folgenden Minimierungsproblems ist: $$\min_{\hat{\boldsymbol{\beta}}} \sum_{i=1}^{n} \hat{u}_i^2$$ [12 Punkte]

b) Was sind die Normalgleichungen? Erläutern Sie ihre Implikationen. [9 Punkte]

c) Zeigen Sie, dass der KQ-Schätzer erwartungstreu ist. Was bedeutet die Eigenschaft Erwartungstreue? [9 Punkte]

[30 Punkte]

Aufgabe 3

Betrachten Sie folgende Schätzergebnisse zu intergenerationellen Zusammenhängen:

Land	Elastizität
Vater-Sohn	
Dänemark	0.071 [0.064, 0.079]
Finnland	0.173 [0.135, 0.211]
Norwegen	0.155 [0.137, 0.174]
Schweden	0.258 [0.234, 0.281]
GB	0.306 [0.242, 0.370]
USA	0.517 [0.444, 0.590]

Anmerkung: 95% Konfidenzintervall in eckigen Klammern.
Quelle: Black und Devereux (2011), Tabelle 1.

a) Welches Regressionsmodell wurde geschätzt? Geben Sie auch an wie die Variablen skaliert sind. [4 Punkte]

b) Kommentieren Sie die Schätzergebnisse: Wie ist der Wert von 0.517 für die USA und von 0.173 für Finnland zu interpretieren? [3 Punkte]

c) Wovon hängt die Stärke des intergenerationellen Zusammenhangs im Allgemeinen ab? Wie lassen sich die Unterschiede zwischen den Ländern erklären? [8 Punkte]

[15 Punkte]

Aufgabe 4

Am Ende der Klausur finden Sie den Output einer empirischen Analyse mit Stata. In dieser Analyse wird der Einfluss von Austattungsmerkmalen von Wohnungen auf die Höhe der Nettomiete untersucht.

a) Schreiben Sie das geschätzte Regressionsmodell 1 auf. [1 Punkte]

b) Geben Sie die partiellen Effekte für das geschätzte Modell 1 an und interpretieren Sie sie. [10 Punkte]

c) Sind die geschätzten Koeffizienten in Regressionsmodell 1 einzeln getestet statistisch signifikant? Geben Sie die einzelnen Schritte des Hypothesentests und die Teststatistik an. Verwenden Sie als Signifikanzniveau $\alpha = 0.05$. [4 Punkte]

d) Was ist die Null- und Alternativhypothese um zu testen, ob die Variablen zur Wohnfläche gemeinsam einen signifikanten Einfluss auf die Nettomiete haben? Geben Sie den Wert der Teststatistik an. Wie fällt die Testentscheidung aus, wenn Sie als Signifikanzniveau $\alpha = 0.05$ verwenden? [3 Punkte]

e) Wozu wurde Regressionsmodell 2 geschätzt? Beschreiben Sie das zugrundeliegende Testverfahren. Welche Rückschlüsse ziehen Sie aus den Schätzergebnissen für Modell 2 und den dazugehörigen Teststatistiken? Was bedeutet das für die Schätzergebnisse in Modell 1? [12 Punkte]

[30 Punkte]

Stata-Output für Aufgabe 4

```
-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  ex4.log
  log type:  text
 opened on:   16 Jul 2014, 11:37:25

. describe

  obs:         2,053
 vars:            13
 size:        65,696
-------------------------------------------------------------------------------
              storage   display
variable name   type    format      variable label
-------------------------------------------------------------------------------
nm              float   %9.0g       Nettomiete in Euro
wfl             int     %8.0g       Wohnfläche in qm
zh              float   %9.0g       Dummy = 1, wenn Zentralheizung
badextra        byte    %8.0g       Dummy = 1, wenn besondere Austattung im Bad
-------------------------------------------------------------------------------
Sorted by:

.
. *Quadrat der Wohnfläche
. generate wflsq = wfl * wfl

.
. *Interaktion zwischen wfl und badextra
. generate wflbad = wfl * badextra
```

```
. *Regressionsmodell 1
. regress nm zh wfl wflsq badextra wflbad

      Source |       SS       df       MS              Number of obs =    2053
-------------+------------------------------           F(  5,  2047) =  504.08
       Model |  68210455.4     5  13642091.1           Prob > F      =  0.0000
    Residual |    55398120  2047  27063.0777           R-squared     =  0.5518
-------------+------------------------------           Adj R-squared =  0.5507
       Total |   123608575  2052  60238.0972           Root MSE      =  164.51

------------------------------------------------------------------------------
          nm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          zh |   176.5765   13.02933    13.55   0.000     151.0244    202.1286
         wfl |   7.470167   .5867794    12.73   0.000      6.31942    8.620914
       wflsq |  -.0046297   .0036636    -1.26   0.206    -.0118144     .002555
    badextra |   108.4255   45.77079     2.37   0.018     18.66331    198.1876
      wflbad |  -.3184789   .4983765    -0.64   0.523    -1.295857    .6588991
       _cons |  -93.34709   25.50357    -3.66   0.000    -143.3627   -43.33143
------------------------------------------------------------------------------

.
.
. test wfl

 ( 1)  wfl = 0

       F(  1,  2047) =  162.07
            Prob > F =    0.0000

. test wfl wflsq wflbad

 ( 1)  wfl = 0
 ( 2)  wflsq = 0
 ( 3)  wflbad = 0

       F(  3,  2047) =  661.48
            Prob > F =    0.0000
```

```
. predict resid, resid

. generate residsq = resid * resid

. predict nmhat
(option xb assumed; fitted values)

. generate nmhatsq = nmhat * nmhat

.
. *Regressionsmodell 2
. regress residsq nmhat nmhatsq

      Source |       SS       df       MS              Number of obs =    2053
-------------+------------------------------           F(  2,  2050) =  145.75
       Model |  7.1766e+11     2  3.5883e+11           Prob > F      =  0.0000
    Residual |  5.0469e+12  2050  2.4619e+09           R-squared     =  0.1245
-------------+------------------------------           Adj R-squared =  0.1236
       Total |  5.7645e+12  2052  2.8092e+09           Root MSE      =   49617

------------------------------------------------------------------------------
     residsq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nmhat |  -31.03096   26.85446    -1.16   0.248    -83.69583    21.63391
     nmhatsq |   .1033378    .020931     4.94   0.000     .0622895    .1443861
       _cons |   7655.731   8248.399     0.93   0.353    -8520.385    23831.85
------------------------------------------------------------------------------

.
. local lmstat1 = e(N) * e(r2)

. display "lmstat1: `lmstat1'"
lmstat1: 255.588101069908

.
. local lmstat2 = e(r2) * e(r2)

. display "lmstat2: `lmstat2'"
lmstat2: .0154989887818218

.
. local crit = invchi2(e(df_m),0.95)

. display "crit: `crit'"
crit: 5.991464547107979
```

```
.
. log close
      name:  <unnamed>
       log:  ex4.log
  log type:  text
 closed on:   16 Jul 2014, 11:37:27
-------------------------------------------------------------------------------
```