Aufgabe 1: [22]

Sie wollen die Skifahrgewohnheiten von Finnen und Schweizern untersuchen. Dazu haben Sie je 130 Einwohner von Saanen (BE,CH) und Kuusamo (Lappland, FI) nach der Anzahl der gefahrenen Kilometer auf Ski (SKIKM) im vergangenen Jahr, ihrem Geschlecht (FE, Dummy: 0=männlich, 1=weiblich), Alter (AGE), Alter quadriert (Age^2), ihrem Einkommen (Y, monatliches Einkommen in 1000,- Euro kaufkraftbereinigt) sowie ihrer sportlichen Fitness (SP1: unsportlich, SP2: sportlich, SP3: sehr sportlich) befragt. Die separate Schätzung des Modells:

$$SKIKM = \beta_1 + \beta_2 FE + \beta_3 Y + \beta_4 AGE + \beta_5 AGE^2 + \beta_6 AGE \cdot Y + \beta_7 SP1 + \beta_8 SP3 + \varepsilon$$

ergibt die umseitig in den Tabellen 1.1 und 1.2 dargestellten Ergebnisse.

a) Interpretieren Sie den Einfluss sportlicher Fitness auf die gefahrenen Ski-Km in Saanen und Kuusamo statistisch und inhaltlich. [4]

- sportlich (SP2) ist Referenzkategorie; SP1 und SP3 sind in Bezug hierauf zu interpretieren
- in der Schweiz gibt es einen starken positiven Zusammenhang zwischen Fitness und gefahrenen Ski-KM (große Koeffizientenwerte, hoch statistisch signifikant)
- in Kuusamo ist der Zusammenhang nicht so klar (SP1 nicht signifikant; SP3 nur auf 10% Niveau), Koeffizientenwerte kleiner
- sehr sportliche Personen in Saanen (Kuusamo) fahren 46,87 km (16,78 km) mehr Ski als unsportliche

b) Testen Sie mittels Teststatistik oder p-Wert, ob der Effekt des Einkommens (Y) auf SKIKM in Saanen mit dem Alter variiert. Wie ist dieser Alterseffekt zu interpretieren? [4]

- Interaktionseffekt ALTER*Y ist auf 5%-Signifikanzniveau signifikant
- Der Effekt der Alters-Einkommensvariation kommt additiv zum Einkommenseffekt hinzu
- Ja, der Einkommenseffekt variiert mit dem Alter
- Bei einem gegebenem Einkommen erhöht ein Jahr höheres Alter den Einkommenseffekt um 0,5
- Beispiel: Bei einem gegebenem Einkommen von 1000 Euro steigt mit jedem Jahr höheren Alters der Effekt des Einkommens auf SKIKM um 0,5 km, z.B. von 27,85 km eines 20-Jährigen mit 1000 Euro Einkommen auf 28,35 km im nächsten Lebensjahr (hinzu kommt noch der Alterseffekt, der hier nicht betrachtet wird)

c) Berechnen Sie für ein gegebenes Einkommen von 3000,-€ das Alter, in dem die Saaner Bevölkerung am weitesten Ski fährt. [3]

- erste Ableitung der Funktion nach Alter
- für Einkommen 3 einsetzen (3000/1000)
- Gleichung umstellen und AGE_opt= 32 Jahre erhalten

d) Berechnen und interpretieren Sie den (marginalen) Effekt des Einkommens einer 36 Jahre alten Person auf die gefahrenen Ski-Kilometer in Kuusamo. [3]

- erste Ableitung der Funktion nach dem Einkommen und für ALTER=36 einsetzen
- marginaler Einkommenseffekt eines 36-Jährigen = 29,31
- Interpretation: Erhöht sich das Einkommen um 1.000,-Euro (!!!), fährt eine 36-jährige Person aus Kuusamo 29,3 km mehr Ski
- (Hinweis: Alters-Einkommenseffekt ist nur auf 10%-Niveau signifikant; zur Bestimmung der Signifikanz des marginalen Effekts müsste die Standardabweichung des marginalen Effekts berechnet werden.)

e) Sie möchten am 1 Prozentniveau testen, ob die Koeffizienten ihrer Ski-Funktion in Finnland und der Schweiz gleich sind. Dazu fügen Sie ihre Stichproben zusammen, schätzen obiges Modell und

erhalten eine Fehlerquadratsumme von 178.000,0. Erläutern Sie Testverfahren und Ergebnis kurz. Wie lauten die Null- und Alternativhypothese? Geben Sie den kritischen Wert der Teststatistik an, den Sie zum Schließen heranziehen. [4]

- Chow-Breakpoint-Test über die Summe der Quadrate
- $SSE_R = 178000;\ SSE_U = 135876 + 35286 = \mathbf{171162}$
- $H_0: \beta_{i,Saanen} = \beta_{i,Kuusamo}$ für i=1,…,8; $H_1: \beta_{i,Saanen} < // > \beta_{i,Kuusamo}$
- $F = \dfrac{(SSE_R - SSE_U)/j}{SSE_U/(T-K)} = \dfrac{(178000-171162)/8}{171162/(260-16)} = 1,33$
- 1,33 < F(1%, 7, 244) = 2,51 → H_0 kann nicht verworfen werden; Koeffizienten in Finnland und in der Schweiz sind gleich (auf 1%-Signifikanzniveau)

f) Führen Sie einen einseitigen Test auf Gleichheit der Fehlertermvarianz der Schätzungen beider Stichproben am 5%-Niveau durch. [4]

- Goldfeld-Quandt-Test
- $H_0: \sigma^2_{Saanen} = \sigma^2_{Kuusamo}$; $H_1: \sigma^2_{Saanen} < \sigma^2_{Kuusamo}$
- $GQ = \sigma^2_{Saanen} / \sigma^2_{Kuusamo} = 33,37^2 / 17,01^2 = 1113,557 / 289,34 = 3,85$

  3,85 > F(5%, 122, 122) = 1,35

→ H_0 wird verworfen; Unterschiedliche Fehlertermvarianz in Saanen und Kuusamo; Heteroskedastie

Tabelle 1.1: Ergebnisse der Schätzung für die Saanen-Stichprobe

Dependent Variable: SKIKM (in Saanen)				
Method: Least Squares				
Included observations:		130		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	85.7525	13.5347	6.3358	0.0000
FE	-18.7500	8.2418	-2.2750	0.0247
Y	17.8593	5.4786	3.2598	0.0014
AGE	19.5000	2.2520	8.6590	0.0000
AGE^2	-0.3281	0.1224	-2.6800	0.0084
AGE * Y	0.5000	0.2518	1.9860	0.0493
SP1	-32.7560	8.8794	-3.6890	0.0003
SP3	46.8720	7.9511	5.8950	0.0000
R-squared	0.4962	Mean dependent var		135.6780
Adjusted R-squared	0.4673	S.D. dependent var		45.7250
S.E. of regression	33.3727	F-Test:		17.1667
Sum squared resid	135876			

Tabelle 1.2: Ergebnisse der Schätzung für die Kuusamo-Stichprobe

Dependent Variable: SKIKM (in Kuusamo)				
Method: Least Squares				
Included observations:		130		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	112.5000	18.7970	5.9850	0.0000
FE	0.1020	0.0617	1.6520	0.1011
Y	12.7500	5.1914	2.4560	0.0155
AGE	22.7900	3.5845	6.3580	0.0000
AGE^2	-0.2850	0.1429	-1.9950	0.0483
AGE * Y	0.4600	0.2558	1.7980	0.0747
SP1	-12.7500	8.7931	-1.4500	0.1496
SP3	16.7801	9.4270	1.7800	0.0776
R-squared	0.5867	Mean dependent var		156.7820
Adjusted R-squared	0.5629	S.D. dependent var		25.7250
S.E. of regression	17.0067	F-Test:		24.7372
Sum squared resid	35286			

Aufgabe 2: [8]

Wir schätzen folgendes Modell: $Y_t = a + \beta_1 X_t + \beta_2 Z_t + e_t$

a) Wie gehen Sie vor, um einen BLUE Schätzer zu erhalten, wenn Heteroskedastie folgender Form vorliegt: $Var(e_t) = s^2(X_t^2 \cdot 4)$ **[4]**

− $Y^* = \dfrac{Y}{2X_t}$ ... oder Transformation $\dfrac{Y}{\sigma 2X_t}$

für alle Variablen

− Transformation notwendig, dann KQ

b) Wie ändert sich die Interpretation der Regressionsparameter nach Schätzung im BLUE Schätzer im Vergleich zum KQ-Schätzer. [2]

− Gar nicht.

c) Das Modell wurde mit 66 Beobachtungen geschätzt. Welchen kritischen F-Wert benötigen Sie bei einem Test auf Gesamtsignifikanz des Modells am 1 Prozentniveau? Geben Sie die relevanten Freiheitsgrade und das α-Niveau an. [2]

$N = 66$; $\alpha=1\%$

$F_{0.99;\,2;\,63} = 4.95$

(exakter Wert ist nicht erforderlich, da in Standardtabellen so nicht gegeben; bei DF=60: 4.98)

Aufgabe 3: (Wahr oder falsch) [6]

Wahr oder Falsch? Tragen Sie für zutreffende Aussagen den Buchstaben w (für wahr), für nicht zutreffende f (für falsche) ein. (Für jede richtige Antwort gibt es 0,5 Punkte, für jede falsche Antwort werden 0,5 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.)

W	Mit einem F-Test kann man die statistische Signifikanz eines geschätzten Koeffizienten

	überprüfen.
W	Zwei einzeln betrachtet statistisch insignifikante Koeffizienten können gemeinsam signifikant sein.
F	Wenn bei einem restringierten KQ Schätzer die Restriktion nicht exakt gilt, bleiben die geschätzten Parameter unverzerrt, aber die Standardfehler werden falsch ausgewiesen.
F	Auch wenn wie im Fall von Heteroskedastie oder Autokorrelation die Standardfehler von geschätzten Koeffizienten nicht korrekt sind, bleibt der t-Test verlässlich.
W	Ausgelassene relevante Variablen führen zu verzerrten KQ Schätzern, wenn sie mit den eingeschlossenen Variablen korreliert sind.
F	Unter Heteroskedastie sind die per KQ geschätzten Standardfehler falsch, aber der Schätzer ist nach wie vor BLUE.
F	Wenn man die Art der vermuteten Autokorrelation in den Fehlertermen nicht kennt, ist es sinnvoll die geschätzten Standardfehler mit Hilfe des White-Schätzers zu korrigieren.
W	Der LM Test auf Autokorrelation erster Ordnung in den Störtermen besteht aus einem Signifikanztest für den geschätzten Koeffizienten des um eine Periode verzögerten Fehlerterms, der zusätzlich ins ursprüngliche Modell aufgenommen wird.
W	Method of Moments Schätzer sind in großen Stichproben konsistent.
W	Mit dem Hausman-Test kann man im konkreten Einzelfall überprüfen, ob ein Instrumentvariablenschätzer einem KQ Schätzer überlegen ist.
F	Ein Instrumentvariablenschätzer ist umso präziser, je geringer die Korrelation mit derjenigen stochastischen erklärenden Variable, die selbst mit dem Störterm korreliert ist.
W	Um das Problem stochastischer erklärender Variablen zu lösen, braucht man für jede stochastische erklärende Variable mindestens eine Instrumentvariable.

Aufgabe 4: (Wahr-falsch-weil) [16]

Sind folgende Aussagen richtig? Erläutern Sie stichwortartig Ihre Auffassung. Bsp.: „Stimmt, weil" bzw. „Stimmt nicht, weil" etc. (Nur bei korrekter Begründung wird die Antwort mit 2 Punkten pro Frage honoriert)

W	**Es ist nicht sinnvoll, zu viele erklärende Variablen im Modell zu berücksichtigen.** → Weil Varianz der geschätzten Koeffizienten steigt.
F	**Der für eine Dummyvariable (0/1) geschätzte Parameter ist identisch und unabhängig davon, "in welche Richtung" die Variable definiert ist, ob zum Beispiel 0 für weiblich und 1 für männlich oder 1 für weiblich und 0 für männlich kodiert ist.** → Er unterscheidet sich im Vorzeichen.
F	**Die Regressionskonstante kann nicht mit erklärenden Variablen korreliert sein.** → z.B. mit vollständigem Satz kategorialer Variablen oder Indikatoren mit geringer Streuung
W	**Bei nicht-stochastischen erklärenden Variablen und Gültigkeit der allgemeinen Annahmen ist der KQ Schätzer auch in großen Stichproben BLUE.** → Wenn in klein, dann auch in groß
W	**Stochastische erklärende Variablen müssen nicht zwingend zu inkonsistenten KQ Schätzern führen.** → Wenn Annahmen $E(e)=0$ und $cov(X,e)=0$ erfüllt oder wenn $E(e\|X)=0$
F	**Jede Variable Z für die gilt, dass $cov(Z,e) = 0$ kann als Instrument dienen, wenn das Problem darin besteht, dass eine erklärende Variable x mit dem Störterm e korreliert ist.** → Bedingung $cov(Z,X) \neq 0$ muss erfüllt sein
F	**Bei einem einseitigen Durbin-Watson Test auf positive Autokorrelation verwerfen wir H0: $\rho=0$ wenn d größer ist als die obere kritische Grenze.**

	→ Wenn $d < d_{LC}$
W	**Bei autokorrelierten Fehlertermen sind Vorhersagen umso genauer, je näher die Vorhersageperiode an der beobachteten Stichprobenperiode liegt.** → Umso mehr Information steckt im letzten beobachteten Störterm für die Vorhersageperiode.

Aufgabe 5: **[3]**

Wir haben einen AR(1) Prozess in den Störtermen unseres Regressionsmodells. Wie lautet $Corr(e_t, e_{t-1})$, wenn $Corr(e_t, e_{t-2}) = 0,25$?

– Da $corr(e_t, e_{t-k}) = \rho^k$ und $\rho^2 = 1/4$ folgt $corr(e_t, e_{t-1}) = \sqrt{1/4} = 1/2$.

Aufgabe 6: [5]

Gegeben sei ein AR(1) Störprozess: $e_t = e_{t-1} + v_t$ für den gilt: $E(e_t) = 0$; $E(v_t) = 0$; $Var(v_t) = s^2{}_v$; $Cov(v_t, v_s) = 0$ für $t \neq s$. Zeigen Sie, dass e_t homoskedastisch ist.

– Homoskedastie, wenn $Var(e_t)$ unabhängig von „t" und konstant für alle „t".

– $Var(e_t) = \sigma_e^2 = Var(\rho e_{t-1} + \upsilon_t)$

$$= Var(\rho e_{t-1}) + Var(\upsilon_t) + 2 \cdot \rho \cdot cov(e_{t-1}, \upsilon_t)$$
$$= \rho^2 \sigma_{e_{t-1}}^2 \qquad + \sigma_\upsilon^2 \qquad + \qquad 0$$

– Wenn $\sigma_{e_t}^2 = \sigma_{e_{t-1}}^2 \Rightarrow \sigma_e^2 = \dfrac{\sigma_\upsilon^2}{1-\rho^2}$

– kein Einfluss von „t" → homoskedastisch