1. Lösungen zu Kapitel 6

Übungsaufgabe 6.1

Aus der Tatsache, dass die einzelnen Koeffizienten der Quartals-Dummies nicht statistisch signifikant von Null verschieden sind, lässt sich nicht die Aussage ableiten, die Saisoneffekte spielten keine Rolle. Es ist möglich, dass keiner der Koeffizienten für sich genommen statistisch signifikant von Null verschieden ist, die Gruppe von Quartals-Dummies jedoch gemeinsam einen Beitrag zur Erklärung des privaten Konsums leisten. Um dies zu testen, lässt sich ein F-Test durchführen, der den gemeinsamen Erklärungsbeitrag der Koeffizienten von Q_1, Q_2 und Q_3 prüft.

Übungsaufgabe 6.2

Das vorliegende Modell kann geschätzt werden. Auch wenn das Modell logarithmierte und quadratische Terme enthält, so ist es doch linear in seinen zu schätzenden Parametern.

Es läge ein Multikollinearitätsproblem vor, wenn Alter und Schulbildung in einem perfekten linearen Verhältnis stehen würden (z.B. $Alter - 6$). Bei erwachsenen Personen, deren Lohn im vorliegenden Beispiel geschätzt wird, ist hiervon jedoch nicht auszugehen.

Darüber hinaus enthält das Modell eine Dummy-Variable, die den Wert 1 für Frauen annimmt. Wird dieses Modell für Männer und Frauen geschätzt, so liegt auch kein so genannter *dummy variable trap* vor. In einem Modell für zwei Gruppen darf man für eine Gruppe einen Dummy einführen.

Übungsaufgabe 6.3

Die allgemeine Testgröße für den F-Test lautet:

$$F = \frac{(ESS_r - ESS_u)/q}{ESS_u/(N-(K+1))}.$$

Durch Umformung kann die Testgröße für die Hypothese, dass alle Koeffizienten mit Ausnahme der Konstanten gleich Null sind, in Abhängigkeit des

Bestimmtheitsmaßes R^2 ausgedrückt werden. Hierfür wird im ersten Schritt das R^2 nach ESS aufgelöst:

$$\begin{aligned} R^2 &= 1 - \frac{ESS}{TSS} \\ \Leftrightarrow ESS &= (1 - R^2)TSS. \end{aligned}$$

In einem Modell, in dem alle Koeffizienten mit Ausnahme der Konstante gleich Null sind, entspricht die *Error Sum of Squares* ($ESS = \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$) der *Total Sum of Squares* ($TSS = \sum_{i=1}^{N}(Y_i - \overline{Y})^2$), da $\hat{Y}_i = \bar{Y}$. Somit ist

$$R_r^2 = 1 - \frac{ESS_r}{TSS_r} = 0.$$

Des Weiteren entspricht die *Total Sum of Squares* (TSS_r) im restringierten Modell der TSS_u im unrestringierten Modell, da beide Modelle dieselbe abhängige Variable Y verwenden. Ersetzt man nun die ESS in der Testgröße ergibt sich:

$$\begin{aligned} F &= \frac{((1 - R_r^2)TSS - (1 - R_u^2)TSS)/q}{(1 - R_u^2)TSS/(N - (K + 1))} \\ &= \frac{((1 - R_r^2) - (1 - R_u^2))TSS/q}{(1 - R_u^2)TSS/(N - (K + 1))} \\ &= \frac{((1 - R_r^2) - (1 - R_u^2))/q}{(1 - R_u^2)/(N - (K + 1))} \\ &= \frac{(R_u^2 - R_r^2)/q}{(1 - R_u^2)/(N - (K + 1))} \\ &= \frac{R_u^2/q}{(1 - R_u^2)/(N - (K + 1))}. \end{aligned}$$

Übungsaufgabe 6.4

a) Bis auf den Koeffizienten der Dummy-Variablen, die den Wert 1 annimmt, wenn das Land eine Demokratie ist, und dem Koeffizienten des quadrierten Maßes für Einkommensungleichheit sind die Koeffizienten einzeln signifikant von Null verschieden.
Beispielhaft soll dies für den Koeffizienten des Bruttosozialprodukts BSP gezeigt werden:

$$|t| = \left|\frac{-0,001}{0,0004}\right| = 2,5 \;\sim\; t_{103}$$

Die t-Statistik überschreitet den kritischen Wert von $1,96$ und die Nullhypothese $H_0 : \beta_1 = 0$ kann mit einer 5%-Irrtumswahrscheinlichkeit verworfen werden.

b) Je höher das Bruttosozialprodukt einer Volkswirtschaft ist, desto geringer ist die durchschnittliche Wachstumsrate des Bruttosozialprodukts pro Kopf. Ein Anstieg des BSP in einem Jahr zwischen 1950 und 1955 um eine Milliarden \$ (in Preisen des Jahres 1950) reduziert die Wachstumsrate der Volkswirtschaft um $0,001$ Prozentpunkte.
   Ebenso reduziert ein hoher Anteil an staatlichen Ausgaben das Wachstum. Steigt die Staatsquote um einen Prozentpunkt, reduziert sich die Wachstumsrate um $0,009$ Prozentpunkte.
   Ein hoher Anteil der Bevölkerung mit einem hohen Bildungsabschluss erhöht hingegen das Wachstum (ein Prozentpunkt Anstieg erhöht die Wachstumsrate um $0,118$ Prozentpunkte).
   Der Demokratie-Dummy ist nicht signifikant von Null verschieden und lässt sich somit nicht interpretieren.
   Schließlich lässt sich noch ein statistisch signifikanter negativer Effekt von Ungleichheit auf die Wachstumsrate erkennen. Erhöht sich das Maß für Einkommensungleichheit um $0,1$, so erhöht sich die Wachstumsrate um $0,091$. Ohne Kenntnis über die Berechnung des zugrundeliegenden Maßes, lässt sich dieser Effekt quantitativ nicht einschätzen.

c) Das R^2 von $0,37$ besagt, dass das Modell 37% der gesamten Varianz in den durchschnittlichen Wachstumsraten der BSP erklären kann.
   Der F-Test testet die gemeinsame Signifikanz aller Koeffizienten. Bei einer Irrtumswahrscheinlichkeit von 5% ist der kritische Wert der F-Verteilung für sechs Zählerfreiheitsgrade und 113 Nennerfreiheitsgrade ca. $2,19$. Da $F > F_{krit}$ kann die Nullhypothese, dass alle Koeffizienten gleich Null sind ($H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$), verworfen werden.

d) Man kann für alle 110 Volkswirtschaften einen eigenen Achsenabschnitt schätzen. Dies ist allerdings nur dann möglich, wenn das Modell ohne Konstante β_0 geschätzt wird. Ansonsten gilt, dass man für i verschiedene Gruppen das Modell mit $i - 1$ verschiedene Dummies schätzen kann.

Übungsaufgabe 6.5

a) Die auf Grundlage des SOEP geschätzte Lohngleichung ergibt einen marginalen Effekt von einem zusätzlichen Jahr Schulausbildung auf die Löhne von Männern von 8%. Dieser Effekt ist statistisch signifikant von Null verschieden ($t = 42 > t_{kritisch}$).
   Arbeitsmarkterfahrung hat ebenfalls einen signifikanten positiven, aber abnehmenden Effekt auf den Lohn. So erhöht ein zusätzliches Jahr Arbeitsmarkterfahrung bei gegebener Anzahl von Jahren an Schulausbildung und gegeben die Anzahl von Zigaretten pro Tag den Lohn um 2,7%. Nach knapp acht Jahren wird dieser Effekt negativ. Während sieben Jahre Berufserfahrung den Lohn noch um 2,1% erhöht, sinkt der Lohn bei acht Jahren Berufserfahrung um 0,8%.
   Frauen haben ein niedrigeres Lohnniveau als Männer. So liegen ihre Löhne im Schnitt um 20,1% unter denen eines Mannes. Darüber hinaus hat ein

Jahr Schulerfahrung einen geringeren Einfluss auf die Löhne von Frauen als auf die Löhne von Männern. Während ein zusätzliches Jahr Schulausbildung die Löhne von Männern wie bereits erwähnt um 8,4% erhöht, beträgt dieser Effekt für Frauen nur 6,9%.

Der Konsum von Zigaretten reduziert die Löhne - von Männern wie von Frauen - jeweils um 2,4% pro Zigarette pro Tag. Auch dieser Effekt ist signifikant von Null verschieden.

b) Mit Hilfe des F-Tests lassen sich verschiedene Ausschlussrestriktionen testen. Die Nullhypothese wäre im vorliegenden Fall $H_0 : \beta_2 = \beta_3 = 0$. Die Testgröße für den F-Test lässt sich aus dem Bestimmtheitsmaß des restringierten (R_r^2) und des unrestringierten Modells (R_u^2) berechnen (s. Aufg. (6.3)).

Das unrestringierte Modell ist das in der Aufgabenstellung geschätzte Modell ($R_u^2 = 0,301$).

Das R_r^2 ergibt sich aus der Schätzung folgender Lohngleichung:

$$\ln(w_i) = \beta_0 + \beta_1 S_i + \beta_2 F_i + \beta_3 F_i \times S_i + \beta_4 R_i + \epsilon_i.$$

Ist der errechnete F-Wert größer als der kritische Wert der F-Verteilung für zwei Zählerfreiheitsgrade und $N - 7$ Nennerfreiheitsgraden, kann die Nullhypothese abgelehnt werden.

c) Es gibt verschiedene Möglichkeiten um zu testen, ob die Dummy-Variablen für die Region einen signifikanten Beitrag zu der Erklärung der Löhne liefern.

Die erste Möglichkeit wäre der zuvor beschriebene F-Test. Hierzu wird das Modell einmal mit und einmal ohne Regionen-Dummies geschätzt und aus den Bestimmtheitsmaßen lässt sich die Testgröße errechnen.

Alternativ lässt sich der Lagrange-Multiplier-Test durchführen. Für diesen Test genügt es im ersten Schritt, das restringierte Modell ohne Regionen-Dummies zu schätzen. Die Nullhypothese lautet, dass die Regionen-Dummies keinen signifikanten Einfluss auf die Löhne ausüben und folglich mit den geschätzten Residuen aus dem restringierten Modell unkorreliert sind.

Im zweiten Schritt werden die Residuen aus dem restringierten Modell ($\hat{\varepsilon}$) auf die erklärenden Variablen aus dem restringierten Modell (in diesem Fall Jahre der Schulausbildung, Berufserfahrung, etc.) und zusätzlich auf die Regionen-Dummies regressiert. Die Testgröße bzw. LM-Statistik berechnet sich als $N \times R_{\hat{\varepsilon}}^2$, wobei q die Anzahl an Restriktionen sprich die Anzahl an Regionen-Dummies ist.

Die Testgröße ist asymptotisch χ_q^2-verteilt. Ist die LM-Statistik größer als der kritische Wert, kann die Nullhypothese verworfen werden.

Übungsaufgabe 6.6

a) Legt man eine 1%-Irrtumswahrscheinlichkeit zugrunde, so ist der kritische Wert der t-Verteilung 2,576. Somit sind der Koeffizient der Dummy-

Variablen, die angibt ob eine Person verheiratet ist, der Koeffizient des quadrierten Alters sowie der Koeffizient des Interaktionsterms zwischen dem Ausländer-Dummy und dem Frauen-Dummy nicht signifikant verschieden von Null.

Bei einer Irrtumswahrscheinlichkeit von 5% wäre lediglich der Koeffizient der Dummy-Variablen, die angibt ob eine Person verheiratet ist, nicht signifikant verschieden von Null.

b) In der geschätzten Lohnregression erhöht ein Jahr zusätzliche Schulausbildung den Bruttomonatslohn signifikant um 8,9%.

Ausländer und Frauen verdienen signifikant weniger als deutsche Männer (15,4% und 20,9%). Für ausländische Frauen ist dieser Effekt noch stärker ausgeprägt, sie verdienen 43,6% weniger als deutsche Männer.

Der Koeffizient der Dummy-Variablen, die angibt ob eine Person verheiratet ist, ist nicht signifikant verschieden von Null und kann daher nicht interpretiert werden.

Das Alter hat einen positiven aber abnehmenden Effekt auf die Löhne. Der marginale Effekt eines zusätzlichen Lebensjahrs hängt folglich vom aktuellen Alter ab. Bei einem Alter von 30 Jahren ist der marginale Effekt 3,7% und bei einem Alter von 45 3,4%.

c) Die Teststatistik für den einseitigen t-Test mit der Nullhypothese $H_0 : \beta_1 = 0,1$ gegen die Alternativhypothese $H_1 : \beta_1 < 0,1$ lautet:

$$t = \frac{0,089 - 0,10}{0,005} = -2,2.$$

Der kritische Wert der t-Verteilung bei einem Signifikanzniveau von 95% ist $-1,645$ und bei einem Signifikanzniveau von 99% $-2,326$. Folglich kann die Nullhypothese mit einer Irrtumswahrscheinlichkeit von 5% abgelehnt werden, nicht jedoch mit einer Irrtumswahrscheinlichkeit von 1%.

d) Das Bestimmtheitsmaß der Schätzung hat den Wert $R^2 = 0,290$, was bedeutet, dass die Variablen 29% der gesamten Variation in den Löhnen erklären können.

Die F-Statistik ist die Testgröße für den Test auf gemeinsame Signifikanz aller Koeffizienten. Der kritische Wert der F-Verteilung bei einer 5%-Irrtumswahrscheinlichkeit ist 2,01. Da $5,45 > 2,01$ kann die Nullhypothese, dass die Koeffizienten gemeinsam insignifikant sind, verworfen werden.

e) Eine Möglichkeit um auf Unterschiede in Ausbildungsrenditen zwischen Männern und Frauen zu testen ist„ einen Interaktionsterm zwischen dem Frauen-Dummy und den Schuljahren zu erstellen.

Anschließend lässt sich mit Hilfe des t-Test die Signifikanz dieses Koeffizienten prüfen.

Übungsaufgabe 6.7

a) In der vorliegenden Form lässt sich Gleichung (6.52) nicht schätzen, da das Modell nicht linear in den Parametern ist. Durch Logarithmieren lässt sich die Produktionsfunktion linearisieren und kann mit einem linearen Regressionsmodell geschätzt werden:

$$\ln Y = \ln A + \alpha \ln L + \beta \ln K.$$

b) Man kann eine Dummy-Variable in das Modell einfügen (D), die den Wert 1 für die Jahre nach der Ölpreiskrise (1973-1990) annimmt.
Dieser Dummy kann mit dem Kapitalbestand K und den Erwerbstätigen L interagiert werden.
Das Modell sieht folgendermaßen aus:

$$\ln Y = \ln A + \alpha_1 \ln L + \alpha_2 (D \times \ln K) + \beta_1 \ln K + \beta_2 (D \times \ln K) + \gamma D.$$

Nun kann getestet werden, ob α_2 und β_2 einzeln und/oder gemeinsam signifikant von Null verschieden sind. Kann die Nullhypothese nicht verworfen werden, so kann man von einer Veränderung der Produktionselastizitäten ausgehen.

Übungsaufgabe 6.8

a) Der Wissenschaftler macht einen Querschnittsvergleich. Er vergleicht den Umsatz zwischen dem Supermarkt Oldi und dem anderer Supermärkte. Der Koeffizient von Interesse ist β_3, welcher den Maßnahmeneffet misst. Formal wird die folgende Identifikationsannehme getroffen:

$$E(U + \Delta | X, O = 1) - E(U | X, O = 1) = \\ E(U + \Delta | X, D = 1) - E(U | X, D = 0),$$

die bedeutet, dass der Umsatz von Oldi ohne Einführung der erhöhten Mehrwertsteuer, den gleichen Umsatz wie die anderen Supermärkte erzielt hätte.
Ein Problem bei diesem Vorgehen ist das mögliche Vorliegen unbeobachtbarer Heterogenität. Umsatzunterschiede zwischen Oldi und anderen Supermärkten könnten durch unbeobachtete Charakteristika wie Warenangebot, Preis, Erreichbarkeit, etc. entstehen.

b) Die Schätzergebnisse ergeben eine Elastizität zwischen Umsatz und Verkaufsfläche von 0,005. Dementsprechend führt eine einprozentige Vergrößerung der Verkaufsfläche zu einer Umsatzsteigerung von 0,005%.
Die Anzahl der Mitarbeiter hat keinen signifikanten Einfluss auf den Umsatz.
Oldi-Supermärkte haben einen signifikant höheren Umsatz als andere Supermärkte. Im Schnitt ist der Umsatz um 18,5% höher.
Nach der Mehrwertsteuererhöhung würde man zunächst mit geringeren Umsätzen für die Oldi-Supermärkte rechnen.

c) Um zu testen, ob $\ln(F_i)$ und P_i gemeinsam signifikant verschieden von Null sind, schätzt man zunächst das restringierte Modell ($\ln(U_i) = \beta_0 + \beta_1 O_i + \varepsilon$).
Anschließend regressiert man die Residuen $\hat{\epsilon}$ auf die ausgelassenen Variablen. Das $R^2_{\hat{\epsilon}}$ verwendet man zur Berechnung der Test-Statistik für den LM-Test (s. Aufg. (6.5)).
d) Durch die zusätzlichen Informationen lässt sich der Differenz-von-Differenzen-Ansatz durchführen.
Man kann die Differenz der Umsätze vor Einführung der Mehrwertsteuererhöhung zwischen Teilnehmern und Nicht-Teilnehmern mit der Differenz der Umsätze nach Einführung der Mehrwertsteuererhöhung vergleichen.
Eine Annahme dieser Identifikationsstrategie ist, dass die Umsatzdifferenz ohne Mehrwertsteuererhöhung zwischen den Supermärkten konstant geblieben wäre.
Es wird das folgende Modell geschätzt:

$$\Delta U_{it} = \beta_1 \Delta F_{it} + \beta_2 \Delta P_{it} + \beta_3 O_{it} + \Delta \varepsilon_{it}$$

Ein Problem würde sich ergeben, wenn es unterschiedliche temporäre Einflüsse auf die beiden Gruppen gäbe. So kann z.B. Oldi zeitgleich mit Einführung der erhöhten Mehrwertsteuer eine große Marketingaktion starten oder eine der beiden Gruppen könnte von einem Lebensmittelskandal betroffen sein. Es könnte auch ein Ashenfelter-Dip vorliegen, wenn die Maßnahme schon vor Juni 2003 bekanntgegeben worden wäre.

Übungsaufgabe 6.9

a) Der Koeffizient α ist signifikant von Null verschieden und hat den Wert 0,01.
Dies bedeutet, dass die Beschäftigung ab 2000 im Schnitt um 0,1% höher lag als für den Zeitraum von 1997-1999.
Eine Identifikationsannahme ist, dass abgesehen von den Effekten der im Modell integrierten abhängigen Charakteristika X_t und der Politikintervention keine anderen Einflüsse einen Effekt auf die Beschäftigung nach 2000 haben dürfen.
b) Auch hier bietet sich der Differenz-von-Differenzen-Ansatz an, um die Beschäftigungseffekte in Region A abzuschätzen.
Der Evaluationsparameter kann geschätzt werden, indem man in das in a) beschriebene Modell zusätzlich einen Dummy für Region A sowie einen Interaktionsterm zwischen dem Dummy D_t und den Dummy für Region A einfügt. Mithilfe des Koeffizienten dieses Interaktionsterms lässt sich der Beschäftigungseffekt abschätzen.
Dies ist jedoch nur solange gültig, wie keine disproportionalen Veränderungen zwischen den Regionen vorliegt, die nicht mit der Politikmaßnahme verbunden ist.