

Ökonometrie Grundlagen, Übung 7

HENRY HAUSTEIN

Aufgabe 1

Kann ich leider zum jetzigen Zeitpunkt nicht beantworten. Ich werde da mal im Forum nachfragen. Idee ist auf jeden Fall eine Lagrange-Optimierung, aber ich schaffe es dabei nicht, alles richtig zusammenzusetzen.

Aufgabe 2

Wir überprüfen wieder, ob $\mathbb{E}(\hat{\beta}_R) = \beta$ ist:

$$\begin{aligned}\mathbb{E}(\hat{\beta}_R) &= \mathbb{E}(\hat{\beta}_{KQ}) - \mathbb{E}((X'X)^{-1}R'(R(X'X)^{-1}R')^{-1} \underbrace{(R\hat{\beta}_{KQ} - r)}_0) \\ &= \mathbb{E}(\hat{\beta}_{KQ}) - \mathbb{E}(0) \\ &= \beta\end{aligned}$$

Das gilt natürlich nur unter der Annahme, dass $R\hat{\beta}_{KQ} = r$ ist.

Aufgabe 3

(a) Diese Restriktion ist gegeben durch

$$\begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = 0$$

(b) Diese Restriktion ist gegeben durch

$$\begin{pmatrix} 0 & 1 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = 0$$

(c) Diese Restriktion ist gegeben durch

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Aufgabe 4

Wir testen

$$H_0 : R\beta = r$$

$$H_1 : \exists \beta_i : R\beta \neq r$$

Die Teststatistik ist gegeben durch

$$F = \frac{\frac{\hat{u}'_R \hat{u}_R - \hat{u}'_{UR} \hat{u}_{UR}}{m}}{\frac{\hat{u}'_{UR} \hat{u}_{UR}}{T-k}} \sim F_{m, T-k}$$

Ist $F > F_{m, T-k; 1-\alpha}$, so wird H_0 abgelehnt.

Aufgabe 5

Ich werde hier aus Gründen der besseren Lesbarkeit und (meiner Meinung nach) besserem Verständnis einige Ideen der linearen Algebra und Analysis aufgreifen

- Wenn man von zwei Spaltenvektoren x und y das Produkt $x'y$ ausrechnen will, so kommt dabei ein Skalar (eine Zahl) raus. Man nennt dieses Produkt deswegen Skalarprodukt und schreibt dafür $\langle x, y \rangle$. Ausgerechnet bedeutet dies $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$.
- Wenn man den Abstand eines Vektors x zum Ursprung wissen möchte, so berechnet man seine Norm $\|x\|$. Konkret rechnet man die euklidische Distanz mittels des Satzes des Pythagoras aus, also $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$. Eine Norm hat unter anderem drei Eigenschaften
 - $\|x\| \geq 0 \quad \forall x$
 - $\|x + y\| \leq \|x\| + \|y\|$
 - $\| \|x\| x \| = \|x\|$

Skalarprodukt und Norm hängen auf eine sehr schöne Weise miteinander zusammen: $\|x\| = \sqrt{\langle x, x \rangle}$ bzw. $\langle x, x \rangle = \|x\|^2$. Mit diesem Wissen können wir den Beweis angehen. Wir starten diesmal mit der zu beweisenden Aussage und formen solange um, bis wir eine offensichtlich wahre Aussage dastehen haben. Da wir nur Äquivalenzumformungen machen werden, muss dann auch die zu beweisende Aussage richtig sein.

$$\begin{aligned}
 \hat{u}'_{UR} \hat{u}_{UR} &\leq \hat{u}'_R \hat{u}_R \\
 \langle \hat{u}_{UR}, \hat{u}_{UR} \rangle &\leq \langle \hat{u}_R, \hat{u}_R \rangle \\
 \|\hat{u}_{UR}\| &\stackrel{*}{\leq} \|\hat{u}_R\| \\
 \|(y - X\hat{\beta}_{UR})'(y - X\hat{\beta}_{UR})\| &\leq \|(y - X\hat{\beta}_R)'(y - X\hat{\beta}_R)\| \\
 \|\langle (y - X\hat{\beta}_{UR}), (y - X\hat{\beta}_{UR}) \rangle\| &\leq \|\langle (y - X\hat{\beta}_R), (y - X\hat{\beta}_R) \rangle\| \\
 \| \|y - X\hat{\beta}_{UR}\| \| &\stackrel{*}{\leq} \| \|y - X\hat{\beta}_R\| \| \\
 \|y - X\hat{\beta}_{UR}\| &\leq \|y - X\hat{\beta}_R\| \\
 &\leq \|y - X(\hat{\beta}_{UR} - (X'X)^{-1}R'[R(X'X)^{-1}R]^{-1}(R\hat{\beta}_{UR} - r))\| \\
 &\leq \|y - X\hat{\beta}_{UR} + X(X'X)^{-1}R'[R(X'X)^{-1}R]^{-1}(R\hat{\beta}_{UR} - r)\| \\
 &\leq \|y - X\hat{\beta}_{UR}\| + \|X(X'X)^{-1}R'[R(X'X)^{-1}R]^{-1}(R\hat{\beta}_{UR} - r)\|
 \end{aligned}$$

Da Normen immer positiv sind, ist die rechte Seite offensichtlich größer als die linke Seite der Ungleichung. Damit muss dann auch $\hat{u}'_{UR} \hat{u}_{UR} \leq \hat{u}'_R \hat{u}_R$ wahr sein. Die einzige Schwäche die ich hier für das Bestimmtheitsmaß sehe, ist, dass es nicht zwischen restringierten und unrestringierten Modellen unterscheidet.

*: Eigentlich müssten bei diesen Normen noch Quadrate stehen, aber da Quadrieren von positiven Zahlen eine monotone Transformation ist, ändern wir nicht die Richtung der Ungleichung.

Aufgabe 6

Wir sollten zuerst den Datensatz vernünftig verarbeitbar machen:

```
1 datensatz = read.csv2("worldagprod.csv", sep = ",")
2 str(datensatz)
3 datensatz[,5] = as.numeric(datensatz[,5])
4 datensatz[,8] = as.numeric(datensatz[,8])
```

- (a) In der ersten Spalte der Variable `datensatz` stehen die Namen der Länder. Diese sind logischerweise nicht benutzbar für eine Regression, weswegen wir sie entfernen:

```
1 modell = lm(output ~ ., data = datensatz[-1])
2 summary(modell)
```

In diesem Modell ist fast kein Parameter signifikant, und selbst wenn, dann nur wenig. Trotzdem hat das Modell einen adjustierten R^2 von 0.9793.

- (b) Wir stellen fest, dass der erste und letzte Eintrag im Datensatz die Ausreißer enthält und entfernen diese:

```
1 plot(datensatz$output)
2 plot(datensatz$workforce, datensatz$output)
3 datensatzReduziert = datensatz[-c(1,22),]
4 modell2 = lm(output ~ ., data = datensatzReduziert[-1])
5 summary(modell2)
```

Diese Veränderung war eher schlecht als recht, jetzt ist nur noch ein Parameter "ein bisschen" signifikant von Null verschieden und der adjustierte R^2 ist auf 0.668 gesunken.

- (c) Wir suchen einfach in den Ergebnissen von (b) nach den Parametern mit dem kleinsten p -Value. Das sind *workforce*, *arable.land* und *fertilizer*

```
1 modell3 = lm(output ~ workforce + arable.land + fertilizer,
2               data = datensatz)
3 summary(modell3)
```

Wir befinden uns auf einem guten Weg: Die Parameter werden wieder signifikant und das adjustierte R^2 steigt auf 0.975. Ein schnellerer Weg um die 3 signifikantesten Parameter zu finden wäre übrigens

```
1 step(modell2)
```

Nun inkludieren wir wieder die Parameter *livestock* und *work.stock*

```
1 modell4 = lm(output ~ workforce + arable.land + fertilizer
2               + livestock, data = datensatz)
3 summary(modell4)
4 modell5 = lm(output ~ workforce + arable.land + fertilizer
5               + livestock + work.stock, data = datensatz)
6 summary(modell5)
```

Wirklich signifikant sind die zusätzlichen Parameter nicht, sie verwässern eher den Einfluss unserer 3 bisherigen Parameter. Aber wenigstens steigt das adjustierte R^2 auf 0.9756 und schließlich auf 0.9801.

(d) Die Cobb-Douglas-Produktionsfunktion ist gegeben durch

$$y = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot x_3^{\beta_3}$$

$$\log(y) = \log(\beta_0) + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \beta_3 \log(x_3)$$

$$\log(y_t) = \log(\beta_0) + \beta_1 \log(x_{1t}) + \beta_2 \log(x_{2t}) + \beta_3 \log(x_{3t}) + u_t$$

```
1  modell16 = lm(log(output) ~ log(workforce) + log(arable.land)
2              + log(fertilizer), data = datensatz)
3  summary(modell16)
4  0.27593 + 0.42580 + 0.28646
```

Addition der Parameter ergibt 0.98819, also fast 1.

(e) Dazu müssen wir das Modell erstmal umformen. Wir nutzen dabei $\beta_3 = 1 - \beta_1 - \beta_2$

$$y = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot x_3^{\beta_3}$$

$$y = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot x_3^{1-\beta_1-\beta_2}$$

$$\log(y) = \log(\beta_0) + \beta_1 \log(x_1) + \beta_2 \log(x_2) + (1 - \beta_1 - \beta_2) \log(x_3)$$

$$= \log(\beta_0) + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \log(x_3) - \beta_1 \log(x_3) - \beta_2 \log(x_3)$$

$$\log(y) - \log(x_3) = \log(\beta_0) + \beta_1 (\log(x_1) - \log(x_3)) + \beta_2 (\log(x_2) - \log(x_3))$$

```
1  modell17 = lm(I(log(output) - log(fertilizer)) ~ I(log(workforce) -
2              log(fertilizer)) + I(log(arable.land) - log(fertilizer)),
3              data = datensatz)
4  summary(modell17)
5  1 - 0.2808 - 0.4280
```

Die Parameter ändern sich recht wenig, $\beta_3 = 0.2912$. Damit der Schätzer unverzerrt ist, muss die Bedingung für die Grundgesamtheit gelten.