Zwischenprüfung

Jan Beck h11814291

13.11.2020

Fallstudie 1

1.1

Die Kleinste-Quadrate-Schätzung (OLS estimate) des Regressionskoeffizienten beträgt 0.07. Er ist so zu interpretieren, wie sich im Schnitt die Pro-Kopf-Ausgaben für öffentliche Schulen (in Dollar pro Kopf) verändert, wenn wir das Pro-Kopf-Einkommen um 1 Dollar pro Kopf erhöhen. 0.07 entspricht 7 Cent.

1.2

Plus/Positiv. Das ist am Vorzeichen des Koeffizienzen Income zu erkennen, der positiv ist und an der steigenden Regressionsgeraden in Figure 1.

1.3

Die Annahme $\mathbb{E}(u \mid X) = 0$ scheint erfüllt zu sein, da etwa gleich viele der Residuen über wie unter der Nulllinie liegen.

Die Annahme der Homoskedaszität scheint verletzt, da sich mit zunehmendem Einkommen die Menge an Residuen die unter der Nullinien liegen erhöht. Sie sind nicht gleichmäßig verteilt für alle Werte die X annimmt.

1.4

a)

```
0.07 * 1000
```

```
## [1] 70
```

b)

```
0.07 * -2300
```

```
## [1] -161
```

1.5

$$Y = \beta_0 \times X^{\beta_1} \times u, \quad \beta_0, u > 0$$

1.6

Die Kleinste-Quadrate-Schätzung (OLS estimate) des Regressionskoeffizienten income beträgt 1.26. Er ist so zu interpretieren, um wieviel Prozent sich im Schnitt die Pro-Kopf-Ausgaben für öffentliche Schulen verändert, wenn wir das Pro-Kopf-Einkommen um 1 Prozent erhöhen. 1.26 entspricht 1.26%. Die Ausgaben steigen im Schnitt also um 1.26% für jedes zusätzliche Prozent an Einkommen.

1.7

Die Annahme $\mathbb{E}(u \mid X) = 0$ scheint erfüllt zu sein, da etwa gleich viele der Residuen über wie unter der Nulllinie liegen.

Die Annahme der Homoskedaszität scheint erfüllt, da im Großen und Ganzen alle Residuen gleichmäßig um die Nulllinie für alle Werte von X verteilt liegen. Zu beachten ist jedoch ein Ausreißer für das höchste Einkommen. Die Aussagekraft des Models für sehr hohe Einkommen könnte dadurch eingeschränkt sein.

1.8

Ich würde das Model 2 bevorzugen, da die Interpretation als relative Änderung (Prozent) die Vergleichbarkeit erhöht (z.B. international, mit anderen Währungen).

Außerdem scheint die Annahme der Homoskedaszität weniger stark verletzt.

Jedoch liegt R2 von Model 1 mit 0.59 leicht über dem von Model 2 (R2 = 0.58) Somit wird mehr der Variantion von Y durch die Variation von X erklärt. Aus meiner Sicht wiegt dieser kleine Unterschied jedoch nicht die beiden voran genannten Vorteile auf.

1.9

Das Model deutet darauf hin, dass die Pro-Kopf-Ausgaben für öffentliche Schulen eines Staates der USA stark vom durchschnittliche Pro-Kopf-Einkommen aller Einwohner des jeweiligen Staates abhängt. Näme man an, dass damit mit weniger Ausgaben pro Kopf auch entsprechend weniger Bildung pro Kopf zu Verfügung steht, dann wäre somit der gleiche Zugang zu Bildung eingeschränkt, da die Bildung zufällig abhängig ist vom Geburts- bzw. Wohnort.

Würde man allen Bürgern der USA möglichst gleichen Zugang zu Bildung ermöglichen, dann wäre eine Maßnahme, das Budget der Schulen über die Bundesregierung von den ärmeren Staaten zu den reicheren Staaten umzuverteilen (ähnliche des Länderfinanzausgleichs in Deutschland).

Eine weitere Möglichkeit bestünde darin die Bildung weniger abhänging von den Ausgaben zu machen, z.B. durch Open Access zu Schulmaterial.

Fallstudie 2

2.1

Der OLS estimate für sales betrangt 0.16213. Dies ist so zu interpretieren, um wieviel sich im Schnitt das Gehalt des CEOs verändert, wenn der Umsatz des Unternehmens um 1 % steigt und wir alle anderen Parameter gleich lassen. Eine Umsatzsteigerung um 1% würde somit (ceteris paribus) zu einem 0.16213% höheren Gehalt führen.

2.2

Der OLS estimate des Interzepts beträgt 4.62092. Dieser ist so zu interpretieren, wie hoch im Schnitt das Gehalt des CEOs (in Tsd. USD) liegt, wenn alle anderen Parameter gleich Null sind. Dieser Wert ist mit der Exponentialfunktion zu verwenden und mit 1000 zu multiplizieren um auf den geschätzten Wert in USD zu kommen.

```
exp(4.62092) * 1000
```

```
## [1] 101587.4
```

2.3

i)

Da wir sales ebenfalls logarithmieren, macht es ebenfalls Sinn profits zu logarithmieren, damit durch die Interpretationsmöglichkeit als relative Änderung (in Prozent) die Vergleichbarkeit zwischen den Variablen bzw. zwischen verschiedenen Branchen (Profite sind stark branchenabhängig) und international (Währungen) gegeben ist.

ii)

0.00004

Interpreation: wie verändert sich im Schnitt ceteris paribus das Gehalt des CEOs in Prozent, wenn die Gewinne um 1 Mio USD steigen. Das Gehalt steigt im Schnitt c.p. um 0.004%, wenn die Gewinne um 1 Mio USD steigen.

2.4

R2 Model 1: 0.29911 R2 Model 2: 0.29934

Das R-Quadrat unterscheidet sich kaum, ist jedoch in Model 2 höher. Im Zweifelsfall würde ich die Variable profits hinzufügen, das das R2 höher ist.

Gemäß Korrelationswerte der Tabelle beträgt die Korrelation zwischen profits und mktval 0.9181280, was durchaus als hoch zu betrachten ist, da der Korrelationskoeffizient nur Werte zwischen -1 und 1 annehmen kann.

Die Einzbeziehung der Variable würde allerdings den Standardfehler des OLS Schätzers für mktval erhöhen, weil dieser definiert ist durch

$$\mathrm{se}(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{n}\,\mathrm{sd}(x_j)\sqrt{1 - R_j^2}}$$

Eine hohe Korrelation mit den anderen Variablen führt c.p. aufgrund von $1 - R_j^2$ zu einem kleineren Wert unter dem Bruchstrich und damit zu einem höheren Standardfehler des Parameters.

Ich würde die Variable profits hinzufügen, da das R2 höher ist und die Korrelation nicht perfekt (=1) ist. Solange der Korrelationskoeffizient zwischen profits und mktval nicht perfekt, sind die Annahmen des MLR nicht verletzt. Der Marktwert errechnet sich zwar oft aus den abgezinsten zukünftigen Gewinnen und ist somit indirekt ermittelbar, jedoch fließen darin auch Erwartungen über zukünftige Profite mit ein und wir erhalten somit diese zusätzliche Informationen durch Hinzunahme der Variable.

2.5

R2 Model 1: 0.29911 R2 Model 3: 0.31815

Das Model 3 hat ein höheres R2 d.h. mehr Varianz von Y wird durch Varianz in X erklärt. Darum würde ich Model 3 wählen.

2.6

```r
5 * 0.01171
```

```
## [1] 0.05855
```

2.7

```r
exp(4.50379 + log(100) * 0.10924 + log(200) * 0.16285 + log(5) * 0.01171) * 1000 # in USD
```

```
## [1] 360876.7
```

2.8

```r
1 - 44.079 / 173
```

```
## [1] 0.7452081
```

2.9

Gesamtanzahl an Jahren im selben Unternehmen wäre linear perfekt abhängig (Verstoß gegen MLR.3), hätte ein CEO ausschließlich Zeit in dem Unternehmen in der Rolle des CEO verbracht (ceoten = Gesamtanzahl an Jahren im Unternehmen).

Es wäre somit keine gute Idee die Variable einzubeziehen.

(Es wäre hingegen unproblematisch die Gesamtanzahl an Jahren die der CEO nicht in der Rolle des CEOs im Unternehmen verbracht hat hinzuzufügen.)

3 Richtig oder Falsch

3.1

1. FALSCH, denn $\beta_1 = -3$. Richtig wäre, wenn X um 2 Einheiten steigt, dann sinkt Y um genau 6 Einheiten.
2. FALSCH, denn Y = -1 + (-3) * 2 = 7. Der Erwartungswert von Y ist 7.

3.2

1. FALSCH, denn

$$\mathrm{V}(\hat{\beta}_1) = \frac{\sigma^2}{N s_x^2}$$

Je größer die Stichprobenvariation s_x^2, desto kleiner ist die Varianz des Schätzers $\hat{\beta}_1$.

2. FALSCH, denn

$$\mathbb{V}\left(\hat{\beta}_0 \mid \boldsymbol{X}\right) = \frac{\sigma^2}{N s_{\boldsymbol{X}}^2} \cdot \frac{1}{N} \sum_{i=1}^{N} X_i^2$$

$\hat{\beta}_0$ ist nicht von X_i^2 abhängig und *kann* somit korrelieren.

3.3

1. RICHTIG
2. FALSCH, das gilt nur für den Erwartungswert von $\hat{\beta}_i$
3. FALSCH, weil MLR.3 Annahme (Mulitkollinearität) verletzt ist. Summe der jeweiligen Stimmen für Partein gleich Summe aller Stimmen.

3.4

1. RICHTIG
2. FALSCH, wir erwarten einen Anstieg von Y um β_2 Prozent.

3.5

FALSCH, da im Sample keine Kinder enthalten waren, sondern nur Erwachsene.