

## Übungsblatt 7 — Multiple Regression

Zu diesem Übungsblatt empfehlen wir neben der Lektüre von Kapitel 6 des Lehrbuches *Introduction to Econometrics* von *Stock & Watson* eine Aufarbeitung mithilfe von Kapitel 6 in unserem Online-Companion [Introduction to Econometrics with R](#).

### Aufgabe 1 — Anwendungsbeispiel: Produktivität im Produzierenden Gewerbe

Nehmen Sie an, dass die durchschnittliche Produktivität von Angestellten im produzierenden Gewerbe (*avgprod*) von zwei Faktoren abhängt: der durchschnittlichen Ausbildungszeit (*avgtrain*) und der durchschnittlichen Begabung (*avgabil*), und das

$$avgprod = \beta_0 + \beta_1 avgtrain + \beta_2 avgabil + u.$$

Die Annahmen des Gauss-Markov-Theorems seien erfüllt. Falls Trainingsprogramme für die Angestellten von Firmen mit unterdurchschnittlicher Begabung verpflichtend angeboten werden (sodass *avgtrain* und *avgabil* negativ korreliert sind), würde dann der Schätzer  $\hat{\beta}_1$ , den man aus einer Regression von *avgprod* auf *avgtrain* erhält, das wahre  $\beta_1$  eher unter- oder überschätzen?

### Aufgabe 2 — Ursachen einer Verzerzten KQ-Schätzung

Was sind mögliche Ursachen einer verzerzten KQ-Schätzung?

- (a) Heteroskedastizität.
- (b) Das Auslassen einer wichtigen Variable.
- (c) Ein Stichproben-Korrelationskoeffizient von 0.95 zwischen zwei erklärenden Variablen im Modell.
- (d) Perfekte Multikollinearität.

Regressor	(1)	(2)	(3)
College ( $X_1$ )	5.46	5.48	5.44
Female ( $X_2$ )	-2.64	-2.62	-2.62
Age ( $X_3$ )		0.29	0.29
Ntheast ( $X_4$ )			0.69
Midwest ( $X_5$ )			0.60
South ( $X_6$ )			-0.27
Intercept	12.69	4.40	3.75
Zusammenf. Statistiken			
$SSR$	157133	154598	153986
$TSS$	190696	190862	191050
$SER$			
$R^2$			
$\bar{R}^2$			
$n$	4000	4000	4000

Tabelle 1: Ergebnisse dreier Regressionen von  $AHE$  auf das Geschlecht und andere Variablen

### Aufgabe 3 — Multiples Regressionsmodell mit Dummyvariablen

Betrachte die Regressionsergebnisse aus der Tabelle 1. Der zugrunde liegende Datensatz umfasst Erhebungen von 4000 Vollzeitbeschäftigten. Diese haben entweder einen High-School- oder einen College-Abschluss und sind zwischen 25 und 34 Jahren alt. Die Variablen sind:

$AHE$  = average hourly earnings (in Dollar)  
 $College$  = Binäre Variable (1 für College, 0 für High-School)  
 $Female$  = Binäre Variable (1 für Frau, 0 sonst)  
 $Age$  = Alter  
 $Ntheast$  = Binäre Variable (1 für Northeast, 0 sonst)  
 $Midwest$  = Binäre Variable (1 für Midwest, 0 sonst)  
 $South$  = Binäre Variable (1 für South, 0 sonst)  
 $West$  = Binäre Variable (1 für West, 0 sonst)

- Berechnen Sie  $\bar{R}^2$  für jede der Regressionen (1), (2) und (3).
- Benutzen Sie die Ergebnisse der Regression (1). Verdienen Beschäftigte mit einem College-Abschluss im Durchschnitt mehr als solche, die nur einen High-School-Abschluss besitzen? Wie viel mehr? Verdienen Männer im Schnitt mehr als Frauen? Wie viel mehr?
- Benutzen Sie die Ergebnisse der Regression (3). Existieren große regionale Differenzen? Warum taucht der Regressor  $West$  nicht in der Regression auf? Juanita ist eine 28-jährige College-Absolventin aus dem Süden. Jennifer ist eine 28-jährige College-Absolventin aus dem mittleren Westen. Wie groß ist der zu erwartende Unterschied ihres Stundenlohns?