

Übungen – Einheit 7:

Aufgabe 7.1:

Über den Zusammenhang zwischen Einkommen und Bildungsniveau wird oft kontroversiell diskutiert. Um diese Beziehung abzubilden wird oft eine sogenannte Mincer-Einkommensgleichung geschätzt (nach dem Ökonom Jacob Mincer). In dieser Einkommensgleichung wird das Einkommen (income) mit den Bildungsjahren (yedu) und die Jahre an Arbeitserfahrung erklärt. Da die Arbeitserfahrung nicht beobachtet wird, nehmen wir stattdessen die „mögliche“ Erfahrung (siehe Punkt a).

- a) Erstellen Sie eine neue Variable, welche die möglichen Arbeitserfahrungsjahre widerspiegelt. Verwenden Sie hierzu die Variablen ybirth (Geburtsjahr) und yedu (Jahre in Bildung) im SOEP Datensatz. Berechnen Sie mögliche Arbeitserfahrungsjahre wie folgt:

$$\text{Erfahrung} = \text{Alter} - \text{yedu} - 6$$

wobei $\text{Alter} = 2009 - \text{ybirth}$ (da die Erhebung 2009 durchgeführt wurde).

(Hinweis: Bei der Erstellung der Variable gehen wir davon aus, dass eine Person mit 6 Jahren schulpflichtig ist, und nach Abschluss der Schulbildung durchgehen erwerbstätig ist.)

Hinweis: Bei der Erstellung der Variable „Erfahrung“ ergibt sich für eine Person eine Erfahrung von -1. Da Arbeitsmarkterfahrung nicht negativ sein kann, löschen wir diese Person von unserem Datensatz.

- b) Schätzen Sie die Mincer-Einkommensgleichung indem Sie das Einkommens (income) durch die Jahre in Ausbildung und mögliche Arbeitserfahrung erklären! Verwenden Sie dazu nur jene Beobachtungen (Personen), die überhaupt ein Einkommen erzielen (d.h. $\text{income} > 0$).
- c) Können Sie etwas zur Güte des Modells sagen? Welcher Anteil der Variation der Einkommen wird durch die beiden Variablen Bildung und Erfahrung erklärt?

Von unserem Regressionsoutput können wir ablesen, dass unser R^2 gerundet 0.08 ist. In anderen Worten ca. 8% der Einkommensvariation wird durch unsere erklärenden Variablen Erfahrung sowie Bildung erklärt. Auf der einen Seite lässt dieser niedrige Wert darauf schließen, dass wir andere wichtige Variablen nicht in unser Modell mit aufgenommen haben (z.B. Beruf, Wohnort usw.). Auf der anderen Seite sollten wir das R^2 nicht überinterpretieren, solange wir eine Theorie wie die Mincer-Gleichung mit unseren Daten testen wollen. (Oft ist man weniger an der Modellgüte interessiert, sondern mehr am Einfluss einzelner Variablen – also an den geschätzten Parametern.)

- d) Der Koeffizient der Jahre in Bildung in der Mincer-Einkommensgleichung wird oft als Bildungsrendite bezeichnet. Er gibt an, um wieviel das Einkommen steigt, wenn ein zusätzliches Jahr in Bildung verbracht wird. Interpretieren Sie die Bildungsrendite in Deutschland! Interpretieren Sie die statistische Unsicherheit assoziiert mit Ihrer Schätzung der Bildungsrendite.

In unserer Regression schätzen wir einen Koeffizienten für Jahre in Bildung (yedu) von 4417.72. Das bedeutet, dass für jedes zusätzliche Jahr an Bildung das jährliche Einkommen um 4417.72 Euro steigt.

Unser geschätzter Effekt ist auch höchst signifikant. Aus der Regressionstabelle können wir ablesen, dass das 95% Konfidenzintervall unserer Schätzung [3887.38 ; 4948.07] beträgt. In anderen Worten, würden wir die gleiche Schätzung unendlich mal durch ein Zufallsexperiment durchführen, so würde sich unser hypothetisch geschätzte Bildungsrendite in 95% aller Fälle in diesem Intervall befinden. Die Enden des Intervalls sind auch fern von 0, sodass wir von einer stark signifikanten positiven Bildungsrendite in unseren Daten ausgehen können.

Wir können auch von unser Regressionstabelle eine t-Statistik von 16.33 ablesen. Diese t-Statistik liegt weit über den kritischen Werten von 1.96 (95%) und 2.58 (99%) und lässt drauf schließen, dass die Bildungsrendite signifikant unterschiedlich von 0 ist [zur Erinnerung: die t-Statistik für einen T-Test unter der Nullhypothese, dass unser Koeffizient gleich 0 ist lautet: $(4417.72 - 0)/270.4868 = 16.33$]

- e) Unter Verwendung Ihrer Antworten in b) und c), würden Sie einer Maßnahme ohne Einwand zustimmen, die den Zugang zur Universität für Schulabsolventen erleichtert? (Hinweis: Denken Sie an die kausale Ursachen-Wirkungsbeziehung, die der Einführung einer solchen Maßnahme eigentlich zugrunde liegen sollte).

Wie bereits unter b) erwähnt, fehlen in unserem einfachen Modell wahrscheinlich wichtige Einflussfaktoren. Man könnte sich vorstellen, dass motiviertere Individuen auf der einen Seite länger in der Schule bleiben und einen höheren Abschluss anstreben. Auf der anderen Seite, ist es wahrscheinlich, dass solche Individuen auch in anspruchsvolleren aber höher bezahlten Berufen arbeiten. Falls dies zutrifft, so würden wir die Bildungsrendite in unserem Modell überschätzen. Mit einer entsprechenden Schlussfolgerung sollte man daher vorsichtig sein, da der geschätzte Parameter vermutlich nicht als Wirkungszusammenhang interpretiert werden kann.

- f) Schätzen Sie den selben Zusammenhang wie unter b), ersetzen Sie aber das Einkommen, die Jahre in Ausbildung und die mögliche Arbeitserfahrung durch den (natürlichen) Logarithmus der entsprechenden Werte. Wenn die Werte sowohl der abhängigen als auch unabhängigen Variablen logarithmiert werden, können die geschätzten Parameter als Elastizitäten interpretiert werden. Die geschätzten Parameter geben daher die erwartete Änderung der endogenen Variable (in Prozent) an, wenn sich eine exogene Variable um 1 Prozent erhöht. Versuchen Sie, die geschätzten Koeffizienten entsprechend zu interpretieren!

(Hinweis: Mit dem EXCEL-Befehl LN() kann der natürliche Logarithmus einer Zahl berechnet werden.)

Hinweis: Eine Beobachtung in unseren Daten hat 0 Jahre Erfahrung. Da der Logarithmus von 0 nicht existiert müssen wir diese Beobachtung vor der Schätzung entfernen.

Vom Regressionsoutput sehen wir, dass sich sowohl Vorzeichen als auch Signifikanzniveau nicht stark gegenüber unseren vorherigen Ergebnissen verändert. Wir schätzen immer noch eine starke und höchst signifikante Bildungsrendite. Erhöht sich die Jahre in Bildung bei 1%, so erhöht sich das jährliche Einkommen um ca. 1.57%.

Einige Politiker argumentieren, dass Bildung nicht förderungswürdig ist. Anstelle sollte man mehr Anstrengungen ausüben, Schulabgänger in Ausbildungsberufe zu integrieren.

- a) Verwenden Sie Ihre Schätzergebnisse aus Aufgabe 7.1.b) und berechnen Sie das erwartete Einkommen eines(r) 35-jährigen Arbeitnehmer(in)
 - i) mit 10 Jahren formaler Bildung
 - ii) mit 18 Jahren formaler Bildung
- b) Wie würden sich Ihre Ergebnisse in Vergleich zu a) ändern, wenn Sie anstatt eines(r) 35-jährigen eine(n) 55-jährigen Arbeitnehmer(in) betrachten würden?

Siehe Excel: Sowohl in a) als auch b) haben Individuen mit höherer Bildung einen höheren Erwartungswert des Einkommens. Der erwartete Unterschied in den Einkommen beträgt in beiden Fällen (also für die 35-Jährige und für den 55-Jährigen) 32,057 Euro. Das liegt daran, dass der Zusammenhang zwischen Einkommen einerseits und Bildung und Erfahrung andererseits als linear modelliert wurde. Die Ergebnisse zeigen ganz klar, dass Personen mit höherem Bildungsniveau viel höhere Einkommen beziehen.

- c) Berechnen Sie nochmal das erwartete Einkommen der in a) und b) beschriebenen Personen. Verwenden Sie aber dieses Mal die Schätzergebnisse aus Aufgabe 7.1.f)!
(Hinweis: Da die Schätzung mit logarithmierten Werten durchgeführt wurde, müssen Sie auch die in a) und b) angegebenen (bzw. ableitbaren) Werte für Bildung und Erfahrung logarithmieren, um das erwartete Einkommen berechnen zu können. Das erwartete Einkommen ist dann ebenfalls in logarithmierten Werten angegeben. Mit dem EXCEL-Befehl EXP() können Sie das logarithmierte Einkommen wieder in „Niveaus (d.h. in Euro-Beträge) umwandeln.)

Siehe Excel: In beiden Fällen ist es wieder so, dass die Person mit höherer Bildung ein höheres Einkommen erzielt. Wenn man die erwarteten logarithmierten Einkommen wieder in Niveaus umrechnet, erkennt man, dass die erwarteten Einkommensunterschiede mit höherem Alter sogar noch größer ausfallen. Das liegt daran, weil die Variablen logarithmisch transformiert wurden. (Das führt dazu, dass Unterschiede bei großen Werten kleiner werden. Ein Anstieg der Erfahrung von 1 auf 2 Jahre erhöht die $\ln(\text{Erfahrung})$ um 0.69, ein Anstieg der Erfahrung von 40 auf 41 Jahre erhöht die $\ln(\text{Erfahrung})$ aber lediglich um 0.02.) Dadurch sind die ersten Jahre (z.B. an Arbeitserfahrung) besonders wichtig – was sehr plausibel ist.

- d) Welche Stellung würden Sie zu der Aussage der Politiker nehmen? Würde sich Ihre Meinung ändern, wenn Sie statt des Modells aus Aufgabe 7.1.b) das Modell mit logarithmierten Werten aus Aufgabe 7.1.f) verwenden?

Von unseren Ergebnissen lässt sich nicht ableiten, dass die vorgeschlagene Politik erstrebenswert ist. Allerdings sollte dabei beachtet werden, dass unsere Bildungsrendite möglicherweise überschätzt wird, weil wichtige Variablen im Modell nicht berücksichtigt werden (siehe Aufgabe 7.1e)