

## Weiterführende Fragen der Ökonometrie

### Übungsaufgaben – Blatt 9

Erarbeiten Sie sich den Artikel “Does compulsory school attendance affect schooling and earnings?” von J. D. Angrist und A. B. Krueger (1991, *Quarterly Journal of Economics*, pp. 979-1014) mit Hilfe der folgenden Fragen:

- (1 Punkt) Welche Kinder und Jugendliche sind in den USA typischerweise schulpflichtig?
- (2 Punkte) Erklären Sie mit eigenen Worten, weshalb sich aus diesen Gesetzen ein Zusammenhang zwischen dem Quartal, in dem ein Mann geboren wurde, und der Dauer der Ausbildung vermuten lässt. Ist dieser Zusammenhang für Männer, die in einem College oder auf der Universität waren, ebenfalls zu vermuten?
- (1 Punkt) Ist ein solcher Zusammenhang zwischen der Ausbildungsdauer und dem Quartal, in dem man geboren wurde, auch in Deutschland zu vermuten?
- (3 Punkte) Verwenden Sie den Datensatz `angrist_krueger.txt` und erzeugen Sie eine Graphik ähnlich der in Figure I des Artikels. [Hinweis: Bilden Sie jeweils den Mittelwert über eine Kohorte, d.h. über alle Ausbildungsdauern für im Jahre 1930 im ersten Quartal Geborene usw. Wenden Sie die `tapply()` Funktion an. Achten Sie dabei darauf, dass sich der Index hier zusammensetzt aus `year` und `quarter`.]
- (2 Punkte) Vergleichen Sie Figure I bis III des Artikels. Warum unterscheidet sich der Verlauf der Zeitreihe in Figure III so sehr von den Verläufen in den beiden anderen Abbildungen?
- (1 Punkt) Bestätigen die Graphiken in Figure I bis III Ihre Vermutungen aus (b)?
- (3 Punkte) Im nächsten Schritt werden die Ausbildungsdaten trendbereinigt. Erläutern Sie, wie im Artikel dabei vorgegangen wird und erstellen Sie eine der ersten Abbildung in Figure IV ähnliche Graphik. [Hinweis: Der Befehl `decaverage()` aus dem package ‘`pastecs`’ eignet sich sehr gut für die Trendbereinigung.] Warum entspricht Ihre Graphik am Rand nicht der in Figure IV?
- (3 Punkte) Eliminieren Sie den Trend der Variable `educ` indem Sie die im Datensatz bereits enthaltene Variable `ma_cj` von `educ` subtrahieren. Erzeugen Sie nun Quartalsdummies und führen dann die Regression von der bereinigten `educ` Variable auf die Quartalsdummies in folgender Form

$$(E_{icj} - MA_{cj}) = \alpha + \sum_{j=1}^3 \beta_j Q_{icj} + \epsilon_{icj}, \quad i = 1, \dots, N, \quad c = 1, \dots, 10, \quad j = 1, 2, 3,$$

für die Jahre 1930 bis 1939 durch. Warum stimmt Ihr Ergebnis nur in etwa, bzw. für das dritte Quartal nicht, überein?

- (3 Punkte) Interpretieren Sie Ihre Ergebnisse in Hinblick auf die Fragestellung, ob die Ausbildungsdauer davon beeinflusst wird, in welchem Quartal man geboren wird. Ändern sich die Ergebnisse, wenn lediglich die Männer betrachtet werden, die nicht von den Gesetzen zur Schulpflicht betroffen waren, die also auch nach Ende der Schulpflicht weiter in Ausbildung waren (Table 1, 1930-1939)?

- (j) (3 Punkte) Welche Methode wird für die Daten, die zur Erstellung von Table II verwendet wurden, angewandt? Beschreiben Sie, wie dabei die einzelnen Gruppen unterschieden werden. Welches Ergebnis ergibt sich? Sind Unterschiede über die Jahre erkennbar? Wie groß sind die Effekte einzustufen?
- (k) (1 Punkt) In den folgenden Schritten wird nun der cp-Effekt, den ein zusätzliches Jahr an Ausbildung auf den Lohn hat, geschätzt. Erstellen Sie zunächst eine Graphik ähnlich der in Figure V. Beachten Sie, dass Ihnen nur die Daten bis 1939 zur Verfügung stehen.
- (l) (3 Punkte) Beschreiben Sie den Verlauf des Graphen in Figure V über die Jahre, ohne auf die einzelnen Quartale einzugehen, und nennen Sie Gründe für dieses Verhalten. Warum wird also für die folgende Untersuchung die Stichprobe der Männer, die zwischen 1930 und 1939 geboren wurden, verwendet? Wie verhält sich der Verlauf über die einzelnen Quartale?
- (m) (2 Punkte) Erläutern Sie den “Wald-Schätzer”, der u.a. für Table III verwendet wird. Benutzen Sie dazu die Notationen aus Problem 15.3 Wooldridge 3e & 4e. Macht es einen Unterschied, wenn Sie die Variable  $z$  dabei umgekehrt kodieren?
- (n) (4 Punkte) Schätzen Sie Table III für 1930-1939 nach und interpretieren Sie die Ergebnisse. [Hinweis: Zeigen und verwenden Sie für die Standardfehler der Differenzen:  $\text{Var}(\overline{lwage}_1 - \overline{lwage}_0) = \frac{\text{Var}(lwage_1)}{\#\{q1=1\}} + \frac{\text{Var}(lwage_0)}{\#\{q1=0\}}$ , wobei  $lwage_1$  die Daten aus dem ersten Quartal und  $lwage_0$  die der anderen enthält, für  $lwage$  iid. Analoges gilt für  $educ$ .] Was sind die Voraussetzungen, damit die Wald-Schätzung korrekt ist? Überprüfen Sie die, die leichter zu zeigen ist.
- (o) (2 Punkte) Zeigen Sie, dass der Unterschied zwischen “Wald”- und OLS-Schätzung in Table III für 1930-1939 nicht statistisch signifikant ist. [Hinweis: Verwenden Sie den Endogenitätstest aus Wooldridge 3e & 4e.]
- (p) (5 Punkte) Ziel dieser Teilaufgabe ist es, die Spalten (1) bis (4) von Table V nachzuschätzen. Erstellen Sie dazu die Variable  $age$ , die das Alter der Männer am 1. April 1980 quartalsgenau enthält und schätzen Sie anschließend die vier Modelle. Verwenden Sie für die TSLS-Spalten IV-Schätzungen, um korrekte Standardfehler zu erhalten. Interpretieren und vergleichen Sie die jeweiligen Koeffizienten zu  $educ$ . Zu welchem Schluss führen die  $\chi^2$ -Werte?
- (q) (2 Punkte) Wie ändern sich die Interpretationen für Spalten (5) bis (8) in Table V?
- (r) (4 Punkte) Für Table VII wurden in die Schätzungen zusätzlich zu den Variablen aus Table V noch Dummies für den Geburtsort aufgenommen. Schätzen Sie Spalten (1) bis (4) von Table VII nach und interpretieren Sie die Ergebnisse analog zu Teil (p).
- (s) (3 Punkte) Damit das Quartal, in dem ein Mann geboren wurde, eine zulässige IV für  $educ$  sein kann, muss unter anderem die Bedingung erfüllt sein, dass die potentielle IV mit dem Fehler der strukturellen Gleichung unkorreliert ist. Wie ist diese Annahme einzustufen?
- (t) (3 Punkte) Fassen Sie die Aussagen des Artikels in etwa 10 Zeilen zusammen.