Lennart Oelschläger (lennart.oelschlaeger@uni-bielefeld.de)
310214 Praktische Übung zu "Einführung in die Mikroökonometrie"
Wintersemester 2023/2024

Aufgabenblatt 1

Lösungen

Überblick

In der gepoolten Querschnittsanalyse liegen mehrere Zufallsstichproben vor, die die gleiche Struktur, d.h. Beobachtungen über die gleichen Variablen aufweisen. Eine Idee ist dann, die Daten in einem *Pool* zusammenzuführen um so eine Stichprobe mit größerem Umfang zu erhalten.

Dabei können verschiedene Probleme entstehen, da sich die Verteilungen der Beobachtungen in den einzelnen Stichproben unterscheiden können. Ein konkretes Problem könnte sein, dass die Stichproben zu unterschiedlichen Zeitpunkten erhoben wurden, zu denen zwar das zugrundeliegende Modell gleichgeblieben ist, die Koeffizienten sich jedoch verändert haben. Verschiedene Problemlagen sollen in diesem und dem nächsten Übungsblatt analysiert werden.

Aufgaben

In dieser Aufgabe soll der Einfluss der Ausbildungsjahre auf den Stundenlohn untersucht werden. Dazu wird das folgende Modell für die Population spezifiziert,

$$wage = \beta_0 + \beta_1 educ + u \tag{1}$$

wobei *wage* den Stundenlohn bezeichnet und *educ* die Ausbildung in Jahren.

1. Unter welcher Annahme ist das obige Modell korrekt spezifiziert?

```
R> data <- wooldridge::cps78_85
R> wage <- exp(data$lwage)
R>
R> ### wage
R> model <- lm(wage ~ educ, data)
R> white <- lm(residuals(model) ~ predict(model) + I(predict(model)^2))
R> summary(white)
```

```
#
# Call:
# lm(formula = residuals(model) ~ predict(model) + I(predict(model)^2))
#
# Residuals:
#    Min     1Q Median     3Q    Max
# -9.542 -2.779 -0.881  1.746 36.477
#
# Coefficients:
#                     Estimate Std. Error t value Pr(>|t|)
# (Intercept)          6.33368    1.74393   3.632 0.000295 ***
# predict(model)      -1.81295    0.47226  -3.839 0.000131 ***
# I(predict(model)^2)  0.12378    0.03175   3.899 0.000103 ***
# ---
# Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
#
# Residual standard error: 4.234 on 1081 degrees of freedom
# Multiple R-squared:  0.01387, Adjusted R-squared:  0.01204
# F-statistic: 7.601 on 2 and 1081 DF,  p-value: 0.0005271

R> ### log wage
R> model <- lm(lwage ~ educ, data)
R> white <- lm(residuals(model) ~ predict(model) + I(predict(model)^2))
R> summary(white)

#
# Call:
# lm(formula = residuals(model) ~ predict(model) + I(predict(model)^2))
#
# Residuals:
#      Min       1Q   Median       3Q      Max
# -2.25540 -0.36830 -0.01883  0.35380  1.86287
#
# Coefficients:
#                     Estimate Std. Error t value Pr(>|t|)
# (Intercept)           2.7540     0.9008   3.057  0.00229 **
# predict(model)       -3.0193     0.9764  -3.092  0.00204 **
# I(predict(model)^2)   0.8189     0.2639   3.103  0.00196 **
# ---
# Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
#
# Residual standard error: 0.5078 on 1081 degrees of freedom
# Multiple R-squared:  0.008831,    Adjusted R-squared:  0.006997
# F-statistic: 4.816 on 2 and 1081 DF,  p-value: 0.008277
```

2. Der Datensatz `cps78_85.csv` aus Wooldridge (2009) beinhaltet Beobachtungen aus zwei unabhängigen Stichproben aus den Jahren 1978 und 1985. Lesen Sie den Datensatz ein

und machen Sie sich ein Bild von den Variablen `wage` und `educ`. Unterscheiden Sie auch zwischen den einzelnen Stichproben.

3. Schätzen Sie das gepoolte Regressionsmodell in (1) und interpretieren Sie die Koeffizienten. Sind die Koeffizienten statistisch signifikant?

4. Plotten Sie die Residuen $\hat{u}$ gegen die gefitteten Werte $\hat{y}$. Ist die Varianz der Residuen konstant? Testen Sie mithilfe des Breusch-Pagan-Tests (z.B. Wooldridge 2009: 273), ob die Annahme der Homoskedastie verletzt ist.

5. Welche Auswirkungen hat Heteroskedastie auf die KQ-Schätzung? Welche Möglichkeiten gibt es, mit dem Problem der Heteroskedastie umzugehen?

6. Prüfen Sie im folgenden Modell, ob Heteroskedastie vorliegt.

$$\log(wage) = \beta_0 + \beta_1 educ + u \tag{2}$$

   Sind die Koeffizienten statistisch signifikant? Wie sind die Parameter β_0 und β_1 zu interpretieren?

7. Schätzen Sie nun ein Modell, das für das zweite Jahr eine eigene Konstante zulässt.

$$\log(wage) = \beta_0 + \beta_1 educ + \gamma_0 y85 + u \tag{3}$$

   Welche Interpretation besitzt γ_0? Ist es gerechtfertigt, die Konstante im Modell zu behalten? Argumentieren Sie!

8. Schätzen Sie als nächstes das folgende Modell, das zusätzlich einen Interaktionsterm enthält.

$$\log(wage) = \beta_0 + \beta_1 educ + \gamma_0 y85 + \gamma_1 y85 \cdot educ + u \tag{4}$$

   Welche Interpretation besitzt γ_1? Bleibt die Interpretation von γ_0 aus Aufgabenteil (g) gleich? Sind die Koeffizienten statistisch signifikant?

9. Wie kann getestet werden, ob die Koeffizienten gemeinsam signifikant sind? Führen Sie den Test durch.