UNIVERSITÄT DUISBURG ESSEN

Lehrstuhl für Ökonometrie
Prof. Dr. Christoph Hanck
M.Sc. Karolina Gliszczynska

Probeklausur Methoden der Ökonometrie

A1 [*20 Punkte*]

Betrachten Sie das lineare Regressionsmodell

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u},$$

wobei $\boldsymbol{X}$ $(n \times k)$ vollen Spaltenrang besitzt. Zudem gilt $E(\boldsymbol{u}|\boldsymbol{X}) = \mathbf{0}$ und $\mathrm{Var}(\boldsymbol{u}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n$.

a) Leiten Sie den Kleinste-Quadrate Schätzer $\widehat{\beta}$ von β, sowie dessen Varianz her.

b) Zeigen Sie dass der Kleinste-Quadrate Schätzer unter den Annahmen

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n} \xrightarrow{P} \mathrm{E}(\boldsymbol{X}^T\boldsymbol{X}) =: \boldsymbol{S}_{\boldsymbol{X}^T\boldsymbol{X}} > 0 \quad \text{und} \quad \frac{\boldsymbol{X}^T\boldsymbol{u}}{n} \xrightarrow{P} \mathbf{0}$$

konsistent ist.

c) Nennen Sie zwei weitere Eigenschaften der Projektionsmatrix $\boldsymbol{P}_{\boldsymbol{X}}$ **oder** der Residuenmachermatrix $\boldsymbol{M}_{\boldsymbol{X}}$ und zeigen Sie diese.
[4 *Punkte*]

A2 [*20 Punkte*]

Betrachten Sie das lineare Regressionsmodell mit $E(\boldsymbol{v}|\boldsymbol{X}) = \mathbf{0}$ und definieren Sie $\boldsymbol{X} := (\boldsymbol{X}_1 : \boldsymbol{X}_2)$

$$\boldsymbol{y} = \boldsymbol{X}_1\beta + \boldsymbol{v}, \tag{1}$$

jedoch führen wir die Schätzung für β basierend auf

$$\boldsymbol{y} = \boldsymbol{X}_1\beta + \boldsymbol{X}_2\gamma + \boldsymbol{u} \tag{2}$$

durch. Das bedeutet, dass wir β mit

$$\widehat{\beta}_{\text{OLS,L}} = \{\boldsymbol{X}_1^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_2})\boldsymbol{X}_1\}^{-1}\boldsymbol{X}_1^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_2})\boldsymbol{y}, \tag{3}$$

schätzen, während Modell (1) das wahre Modell ist. Das heißt, dass wir irrelevante Variablen inkludieren (da $\gamma = \mathbf{0}$).

a) Zeigen Sie, dass der Schätzer $\widehat{\beta}_{\text{OLS,L}}$ der langen Regression (2) unverzerrt ist.

b) Zeigen Sie, dass die Varianz-Kovarianzmatrix von $\widehat{\beta}_{\text{OLS,L}}$ gegeben ist durch

$$Var(\widehat{\beta}_{\text{OLS,L}}|\boldsymbol{X}) = \sigma^2\{\boldsymbol{X}_1^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_2})\boldsymbol{X}_1\}^{-1}.$$

`Hinweis:` Machen Sie sich zunutze, dass der Schätzer in Modell (2) unverzerrt ist und der Fehlerterm in Modell (1) $E(\boldsymbol{v}|\boldsymbol{X}) = \mathbf{0}$ erfüllt. Sie können ferner annehmen, dass $Var(\boldsymbol{v}|\boldsymbol{X}) = \sigma^2\boldsymbol{I}$.

Bearbeiten Sie die Teilaufgabe (c) in `R`. Bitte verwenden Sie `set.seed(123)` zu Beginn Ihres Codes.

c) Sei $(\beta, \gamma) = (2,\ 0)^\top$, $\boldsymbol{X}_1 = \boldsymbol{x}_1$ und $\boldsymbol{X}_2 = \boldsymbol{\iota}$, wobei $\boldsymbol{\iota}$ ein Vektor von Einsen ist, in Modell (2).

- Erzeugen Sie
$$y_t = 2x_{1t} + 0 + u_t, \qquad t = 1, \dots, 25,$$
wobei $x_{1t} \overset{u.i.v.}{\sim} \mathcal{U}(0, 1)$ und $u_t \overset{u.i.v.}{\sim} N(0, 0.25)$.
- Schätzen Sie β, indem Sie einmal die komplette Matrix $\boldsymbol{X}$ nutzen ($\widehat{\beta}_{\text{OLS,L}}$) und einmal nur den Vektor $\boldsymbol{x}_1$ ($\widehat{\beta}_{\text{OLS}}$). Speichern Sie beide Ergebnisse.
`Hinweis:` Das Regressionsmodell der langen Regression ist $y_t = \beta \cdot x_{1t} + \gamma + u_t$.
- Wiederholen Sie das Experiment 1000 mal und berechnen Sie die Stichprobenvarianz der beiden Schätzer. Welcher Schätzer für β hat eine geringere Stichprobenvarianz? Begründen Sie, wieso Ihr Ergebnis in der Theorie zu erwarten war.

d) Würde sich etwas ändern, wenn der Regressor $\boldsymbol{x}_1$ einen Mittelwert von Null hätte? Begründen Sie mit Hilfe von Theorie.

A3 [*20 Punkte*]

Betrachten Sie das lineare Regressionsmodell

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u}, \tag{4}$$

wobei $\boldsymbol{X}$ $(n \times k)$ vollen Spaltenrang besitzt.

a) Prof. Schranck, ein angesehener Ökonometrieprofessor, grübelt: „Wir haben oft das Problem zu kleiner Datensätze, wir brauchen einfach mehr Daten." Er hat eine Idee: „Warum kopieren wir nicht einfach unsere Daten $\boldsymbol{y}$ und $\boldsymbol{X}$? Das verdoppelt n und unsere Schätzungen werden besser." Seine Assistentin Frau Schreckmann wendet ein: „Aber das sind doch dann dieselben Punkte im Streudiagramm, das bringt uns nicht entscheidend weiter!" Zeigen Sie, dass diese Idee wenig zielführend in Bezug auf den Kleinste-Quadrate Schätzer ist, und dass dieser sich insbesondere gar nicht ändert. `Hinweis:` Die neue abhängige Variable und die neuen Regressoren haben nach Kopie Zeilendimension $2n$ und lassen sich dann schreiben als

$$\widetilde{\boldsymbol{y}} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{y} \end{pmatrix} \qquad \text{und} \qquad \widetilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{X} \end{pmatrix}.$$

Bearbeiten Sie die folgenden Teilaufgaben in `R`**. Bitte verwenden Sie** `set.seed(123)` **zu Beginn Ihres Codes.**

b) Sei $\boldsymbol{\beta} = (\beta_1, \beta_2)^\mathsf{T}$ und $\boldsymbol{X} = (\boldsymbol{\iota}, \boldsymbol{x}_2)$, wobei $\boldsymbol{\iota}$ ein Vektor von Einsen ist. Schätzen Sie den Kleinste-Quadrate Schätzer $\widehat{\boldsymbol{\beta}}$ von $\boldsymbol{\beta}$ in folgendem Modell

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \qquad t = 1, \dots, 20. \tag{5}$$

Simulieren Sie dazu jeweils 20 Beobachtungen des Modells wobei $x_{2t} \overset{u.i.v.}{\sim} N(0,1)$ und $u_t \overset{u.i.v.}{\sim} N(0,1)$. Zudem sei $\beta_1 = 1$ und $\beta_2 = 2$.

c) Eine weitere Assistentin von Prof. Schranck, Frau Karolinger, hat eine weitere Idee: „Wir könnten aber die kopierten Daten für $\boldsymbol{y}$ und $\boldsymbol{X}$ jeweils mit einer Konstanten c skalieren, dann bekommen wir wirkliche neue Punkte! Damit gehören dann zu kleine Stichproben der Vergangenheit an!"
Die neue abhängige Variable und die neuen Regressoren haben nach Kopie wieder Zeilendimension $2n$ und lassen sich nun schreiben als

$$y^* = \begin{pmatrix} \boldsymbol{y} \\ c\boldsymbol{y} \end{pmatrix} \qquad \text{und} \qquad \boldsymbol{X}^* = \begin{pmatrix} \boldsymbol{X} \\ c\boldsymbol{X} \end{pmatrix}.$$

- Speichern Sie die neue Matrix $\boldsymbol{X}^*$ sowie den neuen Vektor $\boldsymbol{y}^*$, die jeweils die alten Beobachtungen $\boldsymbol{X}$ und $\boldsymbol{y}$ enthalten sowie die zusätzlichen Beobachtungen $c\boldsymbol{X}$ und $c\boldsymbol{y}$. Setzen Sie $c = -3$.

- Schätzen Sie den Kleinste-Quadrate Schätzer $\widehat{\boldsymbol{\beta}}^*$ von $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*)^\mathsf{T}$ im Modell

$$y_t^* = \beta_1^* + \beta_2^* x_{2t}^* + u_t^*, \qquad t = 1, \dots, 40 \tag{6}$$

  und vergleichen Sie $\widehat{\boldsymbol{\beta}}^*$ mit $\widehat{\boldsymbol{\beta}}$ aus Modell (5).

d) Berechnen Sie die *SSR* für beide Regressionen. Multiplizieren Sie die *SSR* von Modell (5) mit $(1 + c^2)$. Was fällt Ihnen auf?

e) Schätzen Sie für die Regresssionen jeweils die Standardabweichung des Koeffizientenschätzers $\widehat{\beta}_2$ und $\widehat{\beta}_2^*$. Überprüfen Sie den folgenden Zusammenhang

$$\sqrt{\mathrm{Var}(\widehat{\beta}_2^* | \boldsymbol{X}^*)} = \sqrt{(n-2)/(2n-2)}\sqrt{\mathrm{Var}(\widehat{\beta}_2 | \boldsymbol{X})}.$$

Prof. Schranck ist begeistert von den Ergebnissen des Experiments: „Die Standardfehler des Koeffizienten verkleinern sich um den Faktor $\sqrt{(n-2)/(2n-2)}$, wenn wir zusätzliche Beobachtungen $c\boldsymbol{X}$ und $c\boldsymbol{y}$ hinzufügen! Damit werden doch größere Teststatistiken möglich!" Kommentieren Sie diese Aussage kritisch.

A4 [*20 Punkte*]

Der Datensatz *cigarette* (bereits in R eingelesen) enthält für 46 amerikanische Bundesstaaten die jeweils logarithmierten Werte für den durchschnittlichen pro-Kopf Verbrauch (von Personen über 16) an Zigarettenschachteln pro Jahr (LNC), den realen Preis pro Schachtel (LNP) und das durchschnittliche reale pro-Kopf-Einkommen (LNY, in 100 US-Dollar). Außerdem ist die Variable `south` enthalten, die den Wert 1 annimmt, wenn es sich um einen Südstaat handelt und 0 sonst.
Betrachten Sie das Modell

$$LNC_t = \beta_1 + \beta_2 LNP_t + \beta_3 LNY_t + u_t; \;\; t = 1, \ldots, 46 \tag{7}$$

und bearbeiten Sie folgende Aufgaben:

a) Schätzen Sie $\boldsymbol{\beta} := (\beta_1, \beta_2, \beta_3)^\mathsf{T}$ in obigem Modell und plotten Sie die Residuen der Regression gegen LNP, sowie die Residuen gegen LNY. Was fällt Ihnen auf? Gibt es Hinweise auf eine Verletzung der gewöhnlichen Annahmen des linearen Regressionsmodells?

b) Testen Sie die $H_0 : \beta_2 = 0$ zum Niveau $\alpha = 0.05$ je einmal unter Verwendung von nicht Heteroskedastie-robusten und Heteroskedastie-robusten Standardfehlern (Typ HC1). Zu welchen Ergebnissen kommen Sie mit Hilfe der Tests?

c) Betrachten Sie nun

$$LNC_t = \beta_1 + \beta_2 LNP_t + v_t; \;\; t = 1, \ldots, 46. \tag{8}$$

Führen Sie die OLS-Regression für Modell (8) getrennt für die Südstaaten und die übrigen Staaten der USA durch. Benutzen Sie dafür jeweils die Datensätze *cigarette_sued* und *cigarette_ohne_sued* (bereits in R eingelesen).

Überprüfen Sie mit dem Chow-Test, ob der Einfluss des Achsenabschnitts und des Preises auf den Zigarettenkonsum in den Südstaaten und den übrigen Staaten gleich ist.

d) Berechnen Sie die Hebelwirkung (Leverage) für Modell (7) (R-Befehl *hatvalues()*). Entfernen Sie alle Beobachtungen, deren Leverages mehr als 3 mal so groß sind wie das arithmetische Mittel über alle Leverages. Schätzen Sie die Koeffizienten des Modells mit dem bereinigten Datensatz erneut. Vergleichen und interpretieren Sie.