Übungsblatt 13

Aufgabe 1

Betrachten Sie den DGP (für $i = 1, \dots 100$):

- $\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix}, \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}\right)$
- $u_i \sim N(0,1)$
- $y_i = 3 + x_{1i} + x_{2i} + u_i$.

Simulieren Sie für diesen DGP 1000 Realisationen. Passen Sie jeweils ein lineares Modell mit OLS an die Daten an und speichern Sie die geschätzten Koeffizienten. Wiederholen Sie das Vorgehen außerdem mit nicht korrelierten Regressoren. Nutzen Sie dann eine graphische Methode, um die Verteilung der geschätzten Koeffizienten zu vergleichen. Was fällt Ihnen auf und welche Rückschlüsse ziehen Sie hieraus? Hinweis: Um multivariat normalverteilte Zufallszahlen zu ziehen, können Sie die Funktion rmvnorm() aus dem Paket mvtnorm benutzen.

Aufgabe 2

Simulieren Sie für $i = 1, \dots 100$ und $k = 1, \dots 20$ wie folgt verteilte Zufallszahlen:

- $x_{ki} \overset{iid}{\sim} \exp(1)$
- $y_i \overset{iid}{\sim} N(20, 1)$

Betrachten Sie das Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots \beta_{20} x_{20i} + u_i$$

Es soll die Frage untersucht werden, dass mindestens ein Regressor einen Einfluss auf y hat. Testen Sie dazu die Hypothese $H_A : \beta_1 = \dots = \beta_{20} = 0$ indem Sie für jeden Koeffizienten individuell die $H_0 : \beta_i = 0$ testen (für $\alpha = 0.05$). Wenn für mindestens einen der so durchgeführten Tests die H_0 abgelehnt werden kann, dann lehnen Sie auch H_A ab.

Wiederholen Sie das Vorgehen 100 mal. Wie oft lehnen Sie H_A ab?

Sehen Sie ein Problem beim hier beschriebenen Vorgehen? Wenn ja, nutzen Sie ein bekanntes, besser geeignetes Verfahren um H_A zu testen.

Aufgabe 3

Betrachten Sie erneut den DGP aus Aufgabe 2. Erzeugen Sie eine Realisation und nutzen Sie diesmal einen geeigneten Test, um die Hypothesen

$$\begin{aligned} H_A: &\quad \beta_1 = \dots = \beta_{20} = 0, \\ H_B: &\quad \beta_1 = \dots = \beta_5 = 0 \text{ und} \\ H_C: &\quad \beta_1 + 2\beta_2 = 0 \text{ und } \beta_4 = \beta_8 \end{aligned}$$

zum Signifikanzniveau $\alpha = 0.05$ an den simulierten Daten zu testen.

Aufgabe 4

1. Erzeugen Sie 500 Beobachtungen für folgenden DGP:
$$x_t \sim \mathcal{U}(0, 10)$$
$$u_t \sim N(0, x_t^2)$$
$$y_t = u_t$$

2. Plotten Sie y gegen x. Was stellen Sie fest?

3. Schätzen Sie mit oben erzeugten Daten das Modell
$$y_t = \beta_0 + \beta_1 x_t + u_t$$
mit OLS. Testen Sie die $H_0 : \beta_1 = 0$ zum Niveau $\alpha = 0.05$ je einmal unter Verwendung von nicht robusten und robusten Standardfehlern (Typ HC1). Zu welchen Ergebnissen kommen die Tests?

4. Führen Sie eine Simulationsstudie für die Tests aus Aufgabenteil 3 durch. Wiederholen Sie dazu das Vorgehen von oben 1000 mal und vergleichen Sie die Verwerfungsraten. Was stellen Sie fest?

5. Führen Sie Simulationsstudie von oben auch für den DGP
$$x_t \sim \mathcal{U}(0, 500)$$
$$u_t \sim N(0, x_t^2)$$
$$y_t = \beta x_t + u_t$$
für jedes $\beta \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ durch. Betrachten Sie jeweils die Verwerfungsraten und interpretieren Sie das Ergebnis.

Aufgabe 5

Betrachten Sie nochmal Aufgabe 4 von Blatt 9. Eine weitere Möglichkeit zur Modelselektion ist, die Residuen gegen die gefitteten Werte zu plotten. Plotten Sie dies für die drei verschieden Modelle. Für welches Modell würden Sie sich nun entscheiden?

Aufgabe 6

Der Old Faithful Geysir im Yellowstone Nationalpark ist ein Naturphänomen. Ungefähr stündlich sprudelt der Geysir mit einer Höhe von bis zu 55 Meter bis zu 32000 Liter Wasser aus. Es scheint so, als ob die Wartezeit bis zur nächsten Eruption von der Dauer der letzten Eruption abhängt. Im Datensatz faithful aus dem AER-Paket befinden sich Daten zur Wartezeit und zur Dauer der Eruptionen. Plotten Sie die Wartezeit gegen die Dauer. Was fällt auf? Führen Sie eine Strukturbruchanalyse durch! Teilen Sie die Daten in zwei Gruppen auf. Eine, in der die Dauer kleiner als 3.25 min und eine, in der die Dauer größer ist. Hinweis: Nutzen Sie den Chow-Test aus der Vorlesung.