Übung zur Vorlesung Wirtschafts- und Sozialstatistik
(Augustin / Petry)

Zweites Übungsblatt zu Teil III

Einfache lineare Regression unter Messfehlern

Untersucht werden soll eine Stichprobe $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ eines latenten bivariaten stochastischen Merkmals (ξ, η), wobei die einfache lineare Regressionsbeziehung

$$\eta_i = \beta_0 + \beta_1 \xi_i + \epsilon_i, \qquad i = 1, \ldots, n\,,$$

mit den üblichen Annahmen an die Fehler ϵ_i angenommen werde.
Beobachtet werden die zugehörigen, (potentiell) fehlerbehafteten Daten $(X_1, Y_1) \ldots (X_n, Y_n)$, wobei für jedes $i = 1, \ldots, n$, gilt:

$$X_i = \xi_i + \delta_i \tag{1}$$

und

$$Y_i = \eta_i + \gamma_i. \tag{2}$$

Dabei seien die Annahmen des Grundmodells der klassischen Testtheorie erfüllt.

Aufgabe 1 Illustrieren Sie sich diese Ausgangssituation und die unten betrachteten Spezialfälle

- durch einfache, kurze Simulationsuntersuchungen sowie
- graphisch, indem Sie in einem Scatterplot die einzelnen Daten durch Realisationen von Messfehler der Form

$$\delta_i = \begin{cases} d & \\ & \text{mit Wsk} \\ -d & \end{cases} \begin{matrix} 0.5 \\ \\ 0.5 \end{matrix} \qquad \text{bzw.} \qquad \gamma_i = \begin{cases} d & \\ & \text{mit Wsk} \\ -d & \end{cases} \begin{matrix} 0.5 \\ \\ 0.5 \end{matrix} \qquad i = 1, \ldots, n$$

  mit festem d „verrauschen“ und dabei d variieren.

Vergleichen Sie für folgende Spezialfälle den auf (ξ_i, η_i), $i = 1, \ldots, n$, beruhenden „Benchmark-Schätzer“ $\widehat{\beta}_1$ für β_1 mit dem „naiven Schätzer“ $\hat{\beta}_1^-$, der aus $\widehat{\beta}_1$ entsteht, indem man einfach ξ_i durch x_i und η_i durch y_i, $i = 1, \ldots, n$, ersetzt:

Aufgabe 2 (Messfehler nur in abhängigen Variable)
Sei in (1) die Variable $\delta_i \equiv 0$, $i = 1, \ldots, n$.

a) Zeigen Sie, dass dann, falls man $\xi_1, \ldots, \xi_n$ als feste Größen auffasst, gilt:

$$\mathbb{E}(\hat{\beta}_1^-) = \mathbb{E}(\hat{\beta}_1). \tag{3}$$

b) Was bedeutet dies inhaltlich?

c) Bestätigen Sie (3) durch Einsetzen von Y_i, $i = 1, \ldots, n$, in die Regressionsgleichung!

Aufgabe 3 (Ausschließlicher Kovariablenmessfehler)
Sei nun umgekehrt in (2) die Variable $\gamma_i \equiv 0$, $i = 1, \ldots, n$. Für große n gilt ungefähr

$$\hat{\beta}_1^- \approx \left(\frac{Var(\xi_i)}{Var(\xi_i) + Var(\delta_i)} \right) \hat{\beta}_1 \ .$$

a) Was bedeutet dies inhaltlich?

b) [Für Haupt- und Nebenfachstudierende ab dem 4. Semester]
Präzisieren Sie obige Aussage und weisen Sie sie nach.

c) Warum lässt sich die Argumentation aus Aufgabe 2c) hier nicht direkt übertragen?