Partitionierte Regression

Das vorherige Beispiel zeigt:

Wenn wir den Einfluss eines bestimmten Regressors auf y ermitteln wollen, müssen wir alle relevanten Regressoren mit aufnehmen.

Nehmen wir als Beispiel das Modell

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \mathbf{u}$$

Falls $\mathbf{x}_1$ und $\mathbf{x}_2$ untereinander korreliert sind, würde eine Schätzung von β_1 im Modell

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \mathbf{u}$$

zu Fehlinformationen führen, da $\mathbf{x}_2$ in $\mathbf{u}$ enthalten wäre und dann $\mathbf{x}_1$ mit $\mathbf{u}$ (über $\mathbf{x}_2$) korreliert wäre, was die strikte Exogenität verletzen würde.

Deswegen braucht man die multiple Regression, die $\mathbf{x}_1$ und $\mathbf{x}_2$ gleichzeitig berücksichtigt.

Partitionierte Regression

Es gibt aber eine Methode, das Ergebnis der multiplen Regression

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \mathbf{u}$$

durch eine Sequenz von Einfachregressionen zu erhalten, nämlich die **partitionierte Regression:**

1. Schätze $\mathbf{y} = \gamma_0 + \gamma_1 \mathbf{x}_2 + \mathbf{r}_y$ und ermittle die Residuen $\widehat{\mathbf{r}}_y$.
2. Schätze $\mathbf{x}_1 = \delta_0 + \delta_1 \mathbf{x}_2 + \mathbf{r}_1$ und ermittle die Residuen $\widehat{\mathbf{r}}_1$.
3. Schätze $\widehat{\mathbf{r}}_y = \alpha_0 + \alpha_1 \widehat{\mathbf{r}}_1 + \mathbf{u}$

Dann gilt:

$$\widehat{\alpha}_1 = \widehat{\beta}_1$$

Partitionierte Regression

Diese Aussage folgt aus dem für die Regressionstheorie fundamentalen **Frisch-Waugh-Theorem**.

Das Einbeziehen von $\mathbf{x}_2$ in das Modell nimmt den Einfluss dieser Variablen auf $\mathbf{y}$ von $\mathbf{x}_1$ weg.

Mit anderen Worten: OLS **kontrolliert** für den Einfluss von $\mathbf{x}_2$ auf den Zusammenhang zwischen $\mathbf{y}$ und $\mathbf{x}_1$. Das Resultat gilt analog auch für mehrere Variablen gemeinsam (also für Gruppen von Variablen).

Deshalb bezeichnet man Regressoren auch als **Kontrollvariablen**.

Das erlaubt die ceteris-paribus - Interpretation der geschätzten Parameter einer multiplen Regression.