Ökonometrie Tutorium

Blatt 3

Carsten Stahl und Fabian Hentrich

WiSe 2024/2025

Exkurs F-Test - Einführung

- F-Test ist ein Spezialfall eines Wald-Tests:

$$W = (R\widehat{\beta} - \mathbf{r})' \left[R(X'X)^{-1}R'\right]^{-1} (R\widehat{\beta} - \mathbf{r}) \sim \chi^2 \tag{1}$$

- Der F-Test prüft die Signifikanz von Gruppenunterschieden in einer linearen Regression.

$$F_{m,n} \stackrel{d}{=} \frac{\chi^2_m/m}{\chi^2_n/n} \tag{2}$$

- Er dient zum testen mehrerer Hypothesen

Exkurs F-Test - Nullhypothese

Die Nullhypothese lautet:

$$H_0 : \mathbf{R}\beta = \mathbf{r} \tag{3}$$

- $\mathbf{R}$: $q \times k$ Matrix der linearen Einschränkungen
- $\mathbf{r}$: $q \times 1$ Vektor der Konstanten

Exkurs F-Test - Beispiel Nullhypothese

$$\begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Problem hier: der $rank(R) \neq q$

Exkurs F-Test - Berechnung der F-Statistik

Die F-Statistik für den Test lautet:

$$F = \frac{(R\widehat{\beta} - \mathbf{r})' \left[R(X'X)^{-1}R'\right]^{-1} (R\widehat{\beta} - \mathbf{r})}{SSR} \cdot \frac{n - (k+1)}{q}$$

- q: Anzahl der Einschränkungen (also $rank(R)$)
- $n - (k+1)$: Freiheitsgrade

Exkurs F-Test - Verteilung der F-Statistik

Unter der Nullhypothese folgt die F-Statistik einer F-Verteilung mit q und $n - k$ Freiheitsgraden:

$$F \sim F(q, n - k) \tag{4}$$

- Überschreitet der berechnete F-Wert den kritischen Wert, wird H_0 verworfen.
- H0 ablehnen, wenn $F > F_{krit}$
- Der F-Test ermöglicht es, den Gesamteffekt mehrerer Variablen zu testen

Exkurs F-Test - Verbindung zum t-Test

- F-Test ist theoretisch eine Verallgemeinerung des t-Tests
- mit dem F-Test ist es auch möglich nur eine Hypothese zu testen ($H0 : \beta = 0 \rightarrow$ R=1 und r=0)
- die entsprechende F-Statistik entspricht dabei der quadrierten t-Statistik

$$F = t^2 \tag{5}$$

Aufgabe 1

a)

Da es sich um ein multiplikatives Modell handelt, müssen wir die log log Spezifikation schätzen:

$$\ln q_i = \beta_0 + \beta_1 \ln n_i + \beta_2 \ln k_i + u_i$$

Es ergeben sich die Hypothesen für eine valide Cobb-Douglas Funktion:

$$H_0 : \beta_1 + \beta_2 = 1 \quad H_1 : \beta_1 + \beta_2 \neq 1$$

Aufgabe 1

a)

Dies ist aber Äquivalent zu:

$$H_0 : \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = 1 \quad H_1 : \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \neq 1$$

Also ist $R = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$.

Aufgabe 1

b)

Die gesuchte Teststatistik ist:

$$F = \frac{\left(R\widehat{\beta} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\widehat{\beta} - \mathbf{r}\right) \Big/ q}{\widehat{\sigma}^2}$$

- Hierbei steht $\hat{\sigma}^2$ für die Varianz der Residuen (oder den quadrierten Standardfehler)
- r kommt aus der allgemeinen Form des F-Tests also $H_0 : R\beta = r$
- q ist die Dimension von r hier also 1

Aufgabe 1

b)

Da wir wissen, dass $R(X'X)^{-1}R'$ ebenfalls ein Skalar sein muss, können wir den Wert auch unten in den Nenner des Bruchs reinschreiben. Zusammen mit $q = 1$ ergibt sich eine wesentlich einfachere Version des F-tests:

$$F = \frac{\left(R\hat{\beta} - r\right)^2}{\hat{\sigma}^2 R(X'X)^{-1}R'}$$

Aufgabe 1

b)

Die Auswertung geschieht teils über den `gretl` output. So wissen wir aus den Einträgen:

- $\hat{\sigma}^2 = 0.478067^2$
- $\hat{\beta}_1 = 0.714847$
- $\hat{\beta}_2 = 0.207570$

Daraus lässt sich $R\hat{\beta} - r$ berechnen:

$$R\hat{\beta} - r = \hat{\beta}_1 + \hat{\beta}_2 - 1 = -0.077583$$

Außerdem wissen wir aus der Aufgabenstellung, dass $R(X'X)^{-1}R' = 0.0012952$.

Aufgabe 1

b)

Insgesamt ergibt sich also:

$$F = \frac{\left(R\hat{\beta} - r\right)^2}{\hat{\sigma}^2 R(X'X)^{-1}R'} = \frac{(-0.077583)^2}{0.478067^2 \cdot 0.0012952} = 20.3338$$

Unter $q = 1$ kann man auch statt dem F-Test einen t-Test verwenden. Unter diesen Bedingungen gilt:

$$F \sim F_{1,569-3}$$
$$F_{(1-\alpha),1,569-3} \approx 5.050$$

Da $F > F_{1-\alpha}$ wird H_0 unter 5 % Signifikanz abgelehnt. Es ist also sehr unwahrscheinlich, dass es sich bei dem Zusammenhang um eine Cobb-Douglas-Funktion handelt.

Aufgabe 1

c)

Unter restringierten Schätzung ergibt sich:

$$\ln q_i = \beta_0 + \alpha \ln n_i + (1 - \alpha) \ln k_i + u_i$$

Wenn man diesen Term weiter umformt, landet man auf:

$$\ln \frac{q_i}{k_i} = \beta_0 + \alpha \ln \frac{n_i}{k_i} + u_i$$

Also eine neue einfach-Regression auf zwischen den relativen Anteilen von Produktionsmenge und Kapital und Arbeitseinsatz und Kapital.

Aufgabe 1

d)

Da der restringierte Fall weniger flexibel ist, muss das R^2 kleiner werden.

Aufgabe 2

a)

- $H_0 : \beta_1 + \beta_2 = 1$ und $\beta_3 = 1$ gegen $H_1 : \beta_1 + \beta_2 \neq 1$ und $\beta_3 \neq 1$.

Aufgabe 2

b)

Wir benötigen

$$F = \frac{(R\widehat{\beta} - \mathbf{r})' \left[R(X'X)^{-1}R'\right]^{-1} (R\widehat{\beta} - \mathbf{r})}{SSR} \cdot \frac{n - (k+1)}{q} \sim F_{q,n-(k+1)}$$

Aufgabe 2

b)

Wegen $SST = 250$ und $R^2 = 0.8$ wissen wir, dass $SSR = (1 - 0.8) \cdot 250 = 50$ ist. Außerdem ist $R\widehat{\beta} - \mathbf{r}$ hier gegeben durch

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0.62 \\ 0.4 \\ 0.5 \\ 1.1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.1 \\ -0.1 \end{pmatrix}$$

Aufgabe 2

b)

Wegen des Hinweises wissen wir, dass

$$R(X'X)^{-1}R' = \begin{pmatrix} 1 & 7 & 7 & 8 \\ 6 & 11 & 5 & 8 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 8 & 14 \\ 8 & 16 \end{pmatrix}$$

ist. Damit ergibt sich

$$\begin{aligned} F &= \frac{\begin{pmatrix} 0.1 & -0.1 \end{pmatrix} \begin{pmatrix} 8 & 14 \\ 8 & 16 \end{pmatrix}^{-1} \begin{pmatrix} 0.1 \\ -0.1 \end{pmatrix}}{SSR} \cdot \frac{n-(k+1)}{q} \\ &= \frac{0.02875}{50} \frac{200}{2} = 0.0575 \end{aligned}$$

Aufgabe 2

c)

- Die Teststatistik ist F-verteilt mit 2 und 200 Freiheitsgraden.

Aufgabe 2

d)

- Das marginale Signifikanzniveau beträgt 94.41%, d.h. die Nullhypothese würde auf jedem kleinerem Signifikanzniveau als diesem nicht abgelehnt werden. Konventionelle Signifikanzniveaus sind 1%, 5% oder 10%, so dass die Nullhypothese nach keinem dieser Kriterien verworfen werden würde.

Aufgabe 3

a)

In diesem Setup ist R eine Matrix und r ein Vektor:

$$H_0 : \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad H_1 : \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \neq \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Aufgabe 3

a)

Die Test-Statistik bleibt unverändert:

$$F = \frac{\left(R\widehat{\beta} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\widehat{\beta} - \mathbf{r}\right) \Big/ q}{\widehat{\sigma}^2}$$

- Aus $SSR = 83$ und $n = 86$ folgt, dass $\hat{\sigma}^2 = \frac{SSR}{n-k-1} = \frac{83}{83} = 1$
- $q = 2$
- $\hat{\beta}_1 = 2$ und $\hat{\beta}_2 = -0.5$ wissen wir aus der Aufgabenstellung
  - Somit ergibt sich:
    $\left(R\hat{\beta} - r\right)' = \begin{pmatrix} 2-1 & -0.5-(-1) \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \end{pmatrix}$

Aufgabe 3

a)

Wir errechnen für die F-Statistik unter Benutzung der gegebenen $(X'X)^{-1}$:

$$
\begin{aligned}
F &= \frac{(R\widehat{\beta} - \mathbf{r})' \left[R(X'X)^{-1}R'\right]^{-1} (R\widehat{\beta} - \mathbf{r})}{\widehat{\sigma}^2} \cdot \frac{1}{q} \\
&= \left[\begin{pmatrix} 1 & 0.5 \end{pmatrix} \begin{pmatrix} 6 & -6 \\ -6 & 12 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} \right] \cdot \frac{1}{2} \\
&= \left[\begin{pmatrix} 1 & 0.5 \end{pmatrix} \begin{pmatrix} 1/3 & 1/6 \\ 1/6 & 1/6 \end{pmatrix} \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} \right] \cdot \frac{1}{2} \\
&= 0.5417 \cdot \frac{1}{2} = 0.2708
\end{aligned}
$$

Aufgabe 3

b)

Der p-Wert bedeutet, dass ein Test mit dem Signifikanzniveau von 76.34% oder höher H_0 ablehnen würde. Alle Tests mit einem niedrigeren Signifikanzniveau würden also H_0 annehmen.

Die Abweichungen der Schätzungen von den Werten -1 und 1 sind folglich nicht signifikant, da unter den konventionellen Signifikanzniveaus (5% oder 1%) nicht abgelehnt wird.

Aufgabe 4

Sie schätzen die folgenden zwei Regressionen:

$$\begin{aligned} \widehat{y}_i &= 5 + 0.2x_i, \quad n = 240, R^2 = 0.23 \\ \widehat{y}_i &= 6 + 0.3x_i + 2.8z_i, \quad n = 240, R^2 = 0.31 \end{aligned}$$

Wie hoch ist der Standardfehler des geschätzten Koeffizienten von z_i (runden Sie auf zwei Nachkommastellen)?

Aufgabe 4 - Lösung

- Hier ist die erste Gleichung eine Restriktion der zweiten (Koeffizient von z_i ist in der ersten Gleichung auf null restringiert).
- daraus folgt: $F = \frac{R^2 - \tilde{R}^2}{1 - R^2} \cdot \frac{n-(k+1)}{q}$
- also $F = \frac{0.31-0.23}{1-0.31} \frac{237}{1} = 27.48$

Aufgabe 4 - Lösung

- einfache Nullrestriktion auf einen Parameter ($F = t^2$ ist also anwendbar)
- $t = \sqrt{27.48} = 5.42$
- der $se(\widehat{coef}(z_i))$ lässt sich nun über die Formel der t-Statistik $(\widehat{coef}(z_i)/se(\widehat{coef}(z_i)))$ berechnen
- gesuchter Standardfehler $se(\widehat{coef}(z_i)) = 2.8/5.42 = 0.52$

Aufgabe 5

a)

Hier handelt es sich um einen Test auf Strukturbruch. Auf die Regressionsgleichung:

$$y_i = \gamma_0 + \gamma_1 x_i^N + \gamma_2 x_i^H + u_{2i}$$

können wir die Hypothesen formulieren:

$$H_0 : \gamma_1 = \gamma_2 \quad H_1 : \gamma_1 \neq \gamma_2$$

Aufgabe 5

b)

Die Teststatistik lautet:

$$F = \frac{R^2 - R^2_{\text{restringiert}}}{1 - R^2} \frac{n - k - 1}{q}$$

- Intuitiv testen wir also darauf, ob unser Modell signifikant besser wird, wenn wir die Information einfügen, dass die Beobachtungen aus verschiedenen Stichproben stammen
- laut Aufgabenstellung ist $R^2 = 0.1411$
- außerdem wurde gegeben dass $R^2_{\text{restringiert}} = 0.1382$
- $q = 1$, da es nur eine Hypothese gibt
- $n - k - 1 = 100 - 2 - 1 = 97$ Da es in dem unrestringierten Fall zwei regressoren gibt.

Aufgabe 5

b)

Einsetzen in die Formel ergibt:

$$F = \frac{0.1411 - 0.382}{1 - 0.1411} \cdot \frac{97}{1} = 0.6652$$

Aufgabe 5

c)

Bei einem p-Wert von 41.71% würden wir die H_0-Hypothese unter 1% und 10% immer noch nicht verwerfen. Das bedeutet, dass sich die beiden Stichproben nicht systematisch auseinander halten lassen.

Aufgabe 6

a)

- $H_0 : \beta_1 = 1$
- $F = \frac{SSR^{restr.} - SSR^{unrestr.}}{SSR^{unrestr.}} \cdot \frac{n-(k+1)}{q} = \frac{120-100}{100} \frac{82-2}{1} = 0.2 \cdot 80 = 16$

Aufgabe 6

b)

Wegen $t = \sqrt{F} = 4$ ist $t = \frac{2-1}{se(\widehat{\beta}_1)} = 4$ und damit $se(\widehat{\beta}_1) = 0.25$.

Aufgabe 6

c)

- R^2 der zweiten Regression ist 0, weil nur auf eine Konstante regressiert wird
- R^2 der unrestringierten Version:
- $F = t^2 = (2/0.25)^2 = 16$
- $F = \frac{R^2}{1-R^2} \frac{n-(k+1)}{q} \Leftrightarrow \frac{R^2}{1-R^2} \frac{82-2}{1} = 16$ folgt $R^2 = 0.1667$