# Econometric Methods (Econometrics I)

## Lecture 4:

## Inference

Prof. Dr. Kai Carstensen

Kiel University

Winter Term 2023/2024

# Outline of this lecture

1. Testing linear restrictions for a single parameter
2. Confidence intervals for a single parameter
3. Testing general linear restrictions
4. Testing exclusion restrictions

Reference: Wooldridge, Chapter 4; Greene, Chapter 5.

# Example: Wage Equation for Married, Working Women

- We are interested in how family factors (number of children) determine the wage a woman achieves.

- To this end, consider a wage equation for married, working women:

$$
\begin{aligned}
\log(wage) \quad &= \quad \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ \\
&\quad\quad + \beta_4 age + \beta_5 kidslt6 + \beta_6 kidsge6 + u
\end{aligned}
$$

  where the last three variables are the woman's age, number of children less than six, and number of children at least six years of age, respectively.

- We are not only interested in point estimates but we would also like to test the following restrictions:
  - no effect of family status: $\beta_5 = 0$ and $\beta_6 = 0$ (individually)
  - no effect of family status: $\beta_5 = 0$ and $\beta_6 = 0$ (jointly)
  - no effect of children's age: $\beta_5 = \beta_6$

- Data and reference: T. A. Mroz (1987), The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions, Econometrica 55, 765-799.

- Stata commands:

```
use "C:\path\mroz.dta", clear

regress lwage exper expersq educ age kidslt6 kidsge6, vce(robust)
```

- Stata output (robust s.e.'s):

```
Linear regression                               Number of obs =      428
                                                F(  6,   421) =    13.78
                                                Prob > F      =   0.0000
                                                R-squared     =   0.1582
                                                Root MSE      =   .66823
```

| lwage | Coef. | Robust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| exper | .039819 | .0152578 | 2.61 | 0.009 | .0098281 | .06981 |
| expersq | -.0007812 | .0004097 | -1.91 | 0.057 | -.0015865 | .0000241 |
| educ | .1078319 | .0136235 | 7.92 | 0.000 | .0810533 | .1346106 |
| age | -.0014653 | .0059351 | -0.25 | 0.805 | -.0131313 | .0102008 |
| kidslt6 | -.0607106 | .1061006 | -0.57 | 0.567 | -.2692635 | .1478424 |
| kidsge6 | -.014591 | .0293505 | -0.50 | 0.619 | -.0722829 | .0431009 |
| _cons | -.4209078 | .3183346 | -1.32 | 0.187 | -1.046631 | .2048154 |

- In the following, we show how to test the hypotheses formulated above. To make them robust to heteroscedasticity, they should be based on robust variance estimators.

# 1. Testing linear restrictions for a single parameter

# The $t$ statistic

- We want to test a hypothesis regarding one of the coefficients estimated by OLS, $\beta_i$:

$$H_0 : \beta_i = \beta_{i,0} \qquad \text{against} \qquad H_1 : \beta_i \neq \beta_{i,0}.$$

- We choose a significance level $= \Pr(H_0 \text{ is rejected} | H_0 \text{ is correct})$ of $\alpha$. This is also called the size of the test.
- The usual test statistic is:

$$t_i = \frac{\hat{\beta}_i - \beta_{i,0}}{\hat{\sigma}_{\hat{\beta}_i}}$$

- We need to find the asymptotic distribution of the $t$ statistic.

# Asymptotic distribution of the $t$ statistic

- Start from the asymptotic distribution

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ (heteroskedasticity) or $\mathbf{V} = \sigma^2\mathbf{A}^{-1}$ (homoskedasticity).

- Then the asymptotic distribution of a single element is

$$\sqrt{N}(\hat{\beta}_i - \beta_i) \xrightarrow{\text{d}} \text{Normal}(0, v_{ii}^2),$$

where $v_{ii}^2$ is the $i$th element of the main diagonal of $\mathbf{V}$.

- Under $H_0 : \beta_i = \beta_{i,0}$ the asymptotic distribution is

$$\sqrt{N}(\hat{\beta}_i - \beta_{i,0}) \xrightarrow{\text{d}} \text{Normal}(0, v_{ii}^2).$$

- Dividing by the standard deviation yields

$$\sqrt{N}\frac{\hat{\beta}_i - \beta_{i,0}}{v_{ii}} \xrightarrow{\text{d}} \text{Normal}(0, 1).$$

# Asymptotic distribution of the $t$ statistic

- To make the $t$ test feasible, we need to estimate $v_{ii}$.
- Both under heteroskedasticity and homoskedasticity, we found "consistent" estimates of the covariance matrix, $\hat{\mathbf{V}} \xrightarrow{\text{p}} \mathbf{V}$.
- Hence, we have a consistent estimate of $\hat{v}_{ii}^2 \xrightarrow{\text{p}} v_{ii}^2$.
- By Slutzky's theorem, $\hat{v}_{ii} \xrightarrow{\text{p}} v_{ii}$.
- Using the asymptotic equivalence lemma (or Cramer's theorem) this implies that

$$\sqrt{N}\frac{\hat{\beta}_i - \beta_{i,0}}{v_{ii}} \qquad \text{and} \qquad \sqrt{N}\frac{\hat{\beta}_i - \beta_{i,0}}{\hat{v}_{ii}}$$

  have the same asymptotic distribution.
- Therefore,

$$t_i = \frac{\hat{\beta}_i - \beta_{i,0}}{\hat{v}_{ii}/\sqrt{N}} \xrightarrow{\text{d}} \text{Normal}(0, 1).$$

# Asymptotic distribution of the $t$ statistic

- A comment on notation.
- Recall that $\hat{v}_{ii}^2$ is the $i$th diagonal element of

$$\hat{\mathbf{V}} \equiv \widehat{\text{Avar}}(\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})).$$

- Also recall that the approximate distribution of $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} \sim \text{Normal}(\boldsymbol{\beta}, \mathbf{V}_{\hat{\boldsymbol{\beta}}}),$$

where $\mathbf{V}_{\hat{\boldsymbol{\beta}}} \equiv \text{Avar}(\hat{\boldsymbol{\beta}}) = \mathbf{V}/N$.

- Let $\hat{\sigma}_{\hat{\beta}_i}^2$ be the $i$th diagonal element of $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$. Then $\hat{\sigma}_{\hat{\beta}_i} = \hat{v}_{ii}/\sqrt{N}$ and the $t$ statistic can equivalently be written as

$$t_i = \frac{\hat{\beta}_i - \beta_{i,0}}{\hat{\sigma}_{\hat{\beta}_i}} \xrightarrow{\text{d}} \text{Normal}(0, 1).$$

- It turns out that $\hat{\sigma}_{\hat{\beta}_i}$ is, in principle, the same estimator as the one derived from exact small-sample theory (that assumes fixed regressors, homoscedasticity and normality).

# Test decision

- In practice (where $N < \infty$), we use the approximate distribution

$$t_i = \frac{\hat{\beta}_i - \beta_{i,0}}{\hat{\sigma}_{\hat{\beta}_i}} \sim \text{Normal}(0, 1).$$

- For a two-sided test, we need an upper and lower critical value.
- Using the symmetry of the normal distribution, we choose $cv_{hi} = -cv_{lo} = z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.
- Or we may compare $|t_i|$ with $cv_{hi}$.
- Test decision: If $|t_i| > cv_{hi} \longrightarrow$ reject $H_0$.
- Remember: size of the test is only approximately correct.

# Example

- Remember the OLS estimates (heteroscedasticity-robust s.e. in brackets):

$$\begin{aligned}\widehat{\log(wage)} &= \underset{(0.318)}{-0.421} + \underset{(0.015)}{0.040}\, exper - \underset{(0.00041)}{0.00078}\, exper^2 + \underset{(0.014)}{0.108}\, educ \\ &\quad \underset{(0.0059)}{-0.0015}\, age - \underset{(0.106)}{0.061}\, kidslt6 - \underset{(0.029)}{0.015}\, kidsge6\end{aligned}$$

- Let us test the two-sided null hypothesis that the number of young children has no effect on their mother's wage, $H_0 : \beta_5 = 0$.
- Choose significance level of 5%.
- Test statistic: $|t_5| = |\hat{\beta}_5/\hat{\sigma}_{\hat{\beta}_5}| = |-0.061/0.106| \approx 0.57$.
- Critical value: $z_{0.975} = 1.96$
- Test decision: $|t_5| < 1.96 \longrightarrow$ do not reject $H_0$.
- Similarly, you can show that $H_0 : \beta_6 = 0$ cannot be rejected.

- Stata command ($F$-test, explanation see below):

```
test kidslt6
```

- Stata output:

```
( 1) kidslt6 = 0

F( 1, 421) = 0.33
Prob > F = 0.5675
```

- In case the numerator degrees of freedom is 1, the absolute $t$-statistic is just the square root of the $F$-statistic. Or directly use the $p$-value given above.

# 2. Confidence intervals for a single parameter

# Review: what is a confidence interval?

- Confidence intervals for single parameters cover the true parameter with a pre-specified probability of $1 - \alpha$.
- This means the following: Assume we could draw a large number of repeated samples from the population. From each sample we construct a confidence interval. Then a fraction of $1 - \alpha$ of the confidence intervals will cover the unknown population parameter.
- This is also called interval estimation.

# Constructing a confidence interval from the $t$ statistic

- To construct a confidence interval for $\beta_i$, we use the approximate distribution

$$t_i = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim \text{Normal}(0,1).$$

- This implies

$$P(-z_{1-\alpha/2} \leq t_i \leq z_{1-\alpha/2}) = 1 - \alpha.$$

- Solving for $\beta_i$ yields the well-known expression

$$P(\hat{\beta}_i - z_{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_i} \leq \beta_i \leq \hat{\beta}_i + z_{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_i}) = 1 - \alpha.$$

- Note once again that the coverage probability holds only approximately.

- Hence, the (approximate) $1-\alpha$ confidence interval for $\beta_i$ is the set

$$CI(\beta_i) = \left[\hat{\beta}_i - z_{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_i},\ \ \hat{\beta}_i + z_{1-\alpha/2}\hat{\sigma}_{\hat{\beta}_i}\right].$$

# Example

- To construct two-sided 90% confidence intervals, recall the OLS estimates (heteroscedasticity-robust s.e. in brackets):

$$\begin{aligned}\widehat{\log(wage)} &= \underset{(0.318)}{-0.421} + \underset{(0.015)}{0.040}\, exper - \underset{(0.00041)}{0.00078}\, exper^2 + \underset{(0.014)}{0.108}\, educ \\ &\quad \underset{(0.0059)}{-0.0015}\, age - \underset{(0.106)}{0.061}\, kidslt6 - \underset{(0.029)}{0.015}\, kidsge6\end{aligned}$$

- Here are the 90% confidence intervals for $\beta_5$ and $\beta_6$.

$$CI(\beta_5) = [-0.061 - 0.106\, z_{0.95},\ \ -0.061 + 0.106\, z_{0.95}] = [-0.236,\ \ 0.114]$$

$$CI(\beta_6) = [-0.015 - 0.029\, z_{0.95},\ \ -0.015 + 0.029\, z_{0.95}] = [-0.063,\ \ 0.034]$$

- In Stata it is more convenient to use the option `level(90)` and read off the results:

```
regress lwage exper ...  kidsge6, vce(robust) level(90)
```

```
Linear regression                               Number of obs =      428
                                                F(  6,   421) =    13.78
                                                Prob > F      =   0.0000
                                                R-squared     =   0.1582
                                                Root MSE      =   .66823
```

| lwage | Coef. | Robust Std. Err. | t | P>\|t\| | [90% Conf. | Interval] |
|---|---|---|---|---|---|---|
| exper | .039819 | .0152578 | 2.61 | 0.009 | .0146668 | .0649712 |
| expersq | -.0007812 | .0004097 | -1.91 | 0.057 | -.0014566 | -.0001059 |
| educ | .1078319 | .0136235 | 7.92 | 0.000 | .0853738 | .1302901 |
| age | -.0014653 | .0059351 | -0.25 | 0.805 | -.0112491 | .0083186 |
| kidslt6 | -.0607106 | .1061006 | -0.57 | 0.567 | -.2356154 | .1141943 |
| kidsge6 | -.014591 | .0293505 | -0.50 | 0.619 | -.0629748 | .0337928 |
| _cons | -.4209078 | .3183346 | -1.32 | 0.187 | -.9456764 | .1038608 |

Note: both intervals include 0.

# 3. Testing general linear restrictions

# Stating general linear restrictions

- To state $r$ general linear restrictions concerning the $K$-dimensional parameter vector $\boldsymbol{\beta}$ we use restriction matrices.
- $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ against $H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$.
- $\mathbf{R}$ is a $r \times K$ matrix of constants and $\mathbf{r}$ is a $r \times 1$ vector of constants.

# Examples for restriction matrices

- Assume we have three parameters $\beta_0$, $\beta_1$, and $\beta_2$.
- We want to test the joint null hypothesis $H_0 : \beta_0 = 1$ and $2\beta_1 = \beta_2$. This can be written in matrix form as

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \end{pmatrix}}_{\mathbf{R}} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{\mathbf{r}}$$

- To test for joint significance of all parameters, we formulate the null hypothesis $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$. This gives rise to

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{R}=\mathbf{I}} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}}_{\mathbf{r}=\mathbf{0}}$$

- Remember that, approximately (as $N$ gets large),

$$\hat{\beta} \sim \text{Normal}(\beta, \mathbf{V}_{\hat{\beta}}).$$

- Hence,

$$\mathrm{E}(\mathbf{R}\hat{\beta}) = \mathbf{R}\,\mathrm{E}(\hat{\beta}) = \mathbf{R}\beta$$

$$\begin{aligned}\mathrm{Var}(\mathbf{R}\hat{\beta}) &= \mathrm{E}[(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)'] = \mathbf{R}\,\mathrm{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']\mathbf{R}' \\ &= \mathbf{R}\,\mathrm{Var}(\hat{\beta})\mathbf{R}' = \mathbf{R}\mathbf{V}_{\hat{\beta}}\mathbf{R}'\end{aligned}$$

and thus

$$\mathbf{R}\hat{\beta} \sim \text{Normal}(\mathbf{R}\beta, \mathbf{R}\mathbf{V}_{\hat{\beta}}\mathbf{R}').$$

- Theorem: Consider the $K$-dimensional random vector $\mathbf{z}$ with $\mathrm{E}(\mathbf{z}) = \boldsymbol{\mu}$ and $\mathrm{Var}(\mathbf{z}) = \boldsymbol{\Sigma}$ that follows a multivariate normal distribution:

$$\mathbf{z} \sim \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Then the quadratic form $(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})$ is distributed as

$$(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) \sim \chi^2_K.$$

- Applying this rule to the $r$-dimensional random vector $\mathbf{R}\hat{\beta}$ with approximate distribution

$$\mathbf{R}\hat{\beta} \sim \mathrm{Normal}(\mathbf{R}\beta, \mathbf{R}\mathbf{V}_{\hat{\beta}}\mathbf{R}')$$

yields the quadratic form

$$(\mathbf{R}\hat{\beta} - \mathbf{R}\beta)'(\mathbf{R}\mathbf{V}_{\hat{\beta}}\mathbf{R}')^{-1}(\mathbf{R}\hat{\beta} - \mathbf{R}\beta) \sim \chi^2_r.$$

- Under $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, the approximate distribution of the quadratic form is

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\mathbf{V}_{\hat{\boldsymbol{\beta}}}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim \chi^2_r.$$

- To make the test feasible, replace $\mathbf{V}_{\hat{\boldsymbol{\beta}}} = \mathbf{V}/N$ by a consistent estimator $\hat{\mathbf{V}}/N$.
- By the asymptotic equivalence lemma (or Cramer's theorem), this leaves the asymptotic distribution unchanged.
- Hence, we use the *Wald* statistic

$$W \equiv (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}(\hat{\mathbf{V}}/N)\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim \chi^2_r.$$

- This holds both under heteroscedasticity (where we use the heteroscedasticity-consistent estimator for $\mathbf{V}$) and under homoscedasticity (where we use the simplified homoscedasticity-only estimator for $\mathbf{V}$).

# Example

- Remember the OLS estimates (heteroscedasticity-robust s.e. in brackets):

$$\begin{aligned}\widehat{\log(wage)} &= \underset{(0.318)}{-0.421} + \underset{(0.015)}{0.040}\, exper - \underset{(0.00041)}{0.00078}\, exper^2 + \underset{(0.014)}{0.108}\, educ \\ &\quad \underset{(0.0059)}{-0.0015}\, age - \underset{(0.106)}{0.061}\, kidslt6 - \underset{(0.029)}{0.015}\, kidsge6\end{aligned}$$

- We want to test the joint null hypothesis $H_0 : \beta_5 = \beta_6 = 0$.
- This can be written in matrix form as

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{\beta} = \mathbf{0}.$$

- To perform this test in Stata use:

```
test kidslt6 kidsge6
```

- When the estimation in Stata is done using heteroscedasticity-robust standard errors, the Wald test will also be heteroscedasticity-robust.

# Example

- Note that Stata rescales the Wald statistic to have an approximate $F$ distribution,

$$F = W/r \xrightarrow{\mathrm{d}} F(r, \infty).$$

- Instead of using this asymptotic distribution, Stata uses the approximate distribution

$$F = W/r \sim F(r, N-K),$$

which would be exact if the disturbances were normal.

- For our example, Stata reports

```
F( 2, 421) = 0.25
Prob > F = 0.7820
```

- Test decision: due to a $p$-value of 0.78, $H_0 : \beta_5 = \beta_6 = 0$ cannot be rejected.
- The 5% critical value of the $F(2, 421)$ distribution is 3.017.
- The 5% critical value of the $F(2, \infty)$ distribution is 2.996.
- Both exceed the $F$ statistic of 0.25.

# Example

- We can also back out the Wald statistic:

$$W = rF = 2 \cdot 0.25 = 0.5.$$

- The 5% critical value of the $\chi^2_2$ distribution is 5.99.
- Test decision: $W < cv \longrightarrow H_0$ cannot be rejected.

# Example

- Let us also test the null hypothesis $H_0 : \beta_5 = \beta_6$.
- This can be written in matrix form as

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \boldsymbol{\beta} = 0.$$

- To perform this test in Stata use:

```
test kidslt6=kidsge6
```

- Stata reports

```
F( 1, 421) = 0.19
Prob > F = 0.6612
```

- Test decision: due to a $p$-value of 0.66, $H_0 : \beta_5 = \beta_6$ cannot be rejected.

# 4. Testing exclusion restrictions

# The Lagrange Multiplier (LM) testing principle

- Exclusion restrictions (and many more) can sometimes more easily be tested using the Lagrange Multiplier (LM) test.
- This test has the advantage that it requires to estimate the model only under the null hypothesis.
- This may be computationally much easier in certain situations than to estimate the model under the alternative as is done for the Wald test.
- More on testing principles later.
- In the following we present (without proof) how to use the LM test to test exclusion restrictions in the linear regression model.

## Testing exclusion restrictions

- Model: $y = \mathbf{x}_1\boldsymbol{\beta}_1 + \mathbf{x}_2\boldsymbol{\beta}_2 + u$
- Hypotheses: $H_0 : \boldsymbol{\beta}_2 = 0$ against $H_1 : \boldsymbol{\beta}_2 \neq 0$
- Step 1: estimate under $H_0$, i.e., estimate by OLS the model

  $$y = \mathbf{x}_1\boldsymbol{\beta}_1 + u,$$

  and compute the residual $\tilde{u} = y - \mathbf{x}_1\tilde{\boldsymbol{\beta}}_1$, where $\tilde{\boldsymbol{\beta}}_1$ is the OLS estimator.
- Step 2: estimate the auxiliary model

  $$\tilde{u} = \mathbf{x}_1\boldsymbol{\alpha}_1 + \mathbf{x}_2\boldsymbol{\alpha}_2 + v$$

  and compute the $R$-squared, $R^2_{\tilde{u}}$.
- Then the LM or score statistic is

  $$LM \equiv NR^2_{\tilde{u}}.$$
- Under $H_0$, $LM$ is $\chi^2_r$ distributed, where $r$ is the number of restrictions.

- In practice, we thus have to run two OLS regressions.
- However, the test statistic $LM \equiv NR^2_{\tilde{u}}$ requires homoscedasticity.
- Wooldridge (p. 59-60) shows how to adjust the auxiliary regression and the test statistic to obtain a heteroscedasticity-robust test.
- For our example, try the LM test in the computer tutorial.