University of Nottingham
UK | CHINA | MALAYSIA

Autumn 2018

# Econometric Theory I

## Computer Lab Class III

Juergen Amann
juergen.amann@nottingham.ac.uk
Wednesday 12:00 - 13:00, C42 SCGB

## Last week

Two exercises:

- **Exercise 1:** Examine the evidence of a gender pay gap.
  - Dummy variables.
  - What does it mean when we *'control for (a) variable(s)'*?
- **Exercise 2:** Examine the determinants of educational attainment.
  - Joint parameter hypothesis testing of coefficients; e.g. $H_0 : \beta_2 = \beta_3$.
  - Restricted Least Squares.

## This week

Model the birth weights of infants:

- Identify factors that may determine a child's birth weight.
- Interpret Stata output.
- One- vs. two-sided hypothesis testing.
- Over- and under-specification of an econometric model.
- Homoskedasticity vs. heteroskedasticity.

## This week

- What are the factors that determine a child's birth weight?

## This week

- What are the factors that determine a child's birth weight?
- Our data set `bwght.xls` contains $n = 1,388$ observations on the following variables:
  - $\texttt{bwght}_{\texttt{i}}$: birth weight in ounces
  - $\texttt{faminc}_{\texttt{i}}$: family income in USD 1,000
  - $\texttt{order}_{\texttt{i}}$: birth order of child
  - $\texttt{cigs}_{\texttt{i}}$: cigarettes smoked per day by mother during pregnancy
  - $\texttt{mothereduc}_{\texttt{i}}$: number of years mother spent in full-time education

## This week

- What are the factors that determine a child's birth weight?
- Our data set bwght.xls contains $n = 1,388$ observations on the following variables:
  - $\texttt{bwght}_\texttt{i}$: birth weight in ounces
  - $\texttt{faminc}_\texttt{i}$: family income in USD 1,000
  - $\texttt{order}_\texttt{i}$: birth order of child
  - $\texttt{cigs}_\texttt{i}$: cigarettes smoked per day by mother during pregnancy
  - $\texttt{mothereduc}_\texttt{i}$: number of years mother spent in full-time education
- What is the expected sign for the variables? Which variables do you expect to be significant?

## Birth weight: regression 1

- Download `bwght.xls` and import it into Stata.

## Birth weight: regression 1

- Download `bwght.xls` and import it into Stata.
- Estimate a regression of birth weight on a constant, family income and birth order.

## Birth weight: regression 1

- Download `bwght.xls` and import it into Stata.
- Estimate a regression of birth weight on a constant, family income and birth order.

```
. regress bwght faminc order
```

## Birth weight: regression 1

- Download `bwght.xls` and import it into Stata.
- Estimate a regression of birth weight on a constant, family income and birth order.

```
. regress bwght faminc order

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |  9103.2286         2  4551.6143   Prob > F        =    0.0000
    Residual | 565508.491     1,385  408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |  574611.72     1,387  414.283864   Root MSE        =    20.207

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      faminc |   .1235904   .0290392     4.26   0.000     .0666249    .1805559
       order |   1.439648   .6086762     2.37   0.018     .2456214    2.633675
       _cons |   112.7618    1.45622    77.43   0.000     109.9052    115.6185
------------------------------------------------------------------------------
```

- Test if variables are significant, interpret signs and magnitudes.

## Birth weight: regression 1

Results of regression 1:

- `_cons`: For `order` = `faminc` = 0, the average expected birth weight is 112.76 ounces, i.e. $\approx 3.18kg$.

## Birth weight: regression 1

Results of regression 1:

- `_cons`: For `order` = `faminc` = 0, the average expected birth weight is 112.76 ounces, i.e. $\approx 3.18kg$.
- `faminc`: Keeping `order` constant, a USD 1,000 increase in family income corresponds to an increase of expected birth weight of 0.12 ounces, i.e. $\approx 3.4g$.

## Birth weight: regression 1

Results of regression 1:

- `_cons`: For `order` = `faminc` = 0, the average expected birth weight is 112.76 ounces, i.e. $\approx 3.18kg$.
- `faminc`: Keeping `order` constant, a USD 1,000 increase in family income corresponds to an increase of expected birth weight of 0.12 ounces, i.e. $\approx 3.4g$.
- `order`: For the same family income, the next-born child is expected to be approx. 1.14 ounces ($\approx 32.4g$) heavier than his/her sibling born right before him/her.

## Birth weight: regression 1

Results of regression 1:

- `_cons`: For `order` = `faminc` = 0, the average expected birth weight is 112.76 ounces, i.e. $\approx 3.18kg$.
- `faminc`: Keeping `order` constant, a USD 1,000 increase in family income corresponds to an increase of expected birth weight of 0.12 ounces, i.e. $\approx 3.4g$.
- `order`: For the same family income, the next-born child is expected to be approx. 1.14 ounces ($\approx 32.4g$) heavier than his/her sibling born right before him/her.

## Birth weight: regression 1

Results of regression 1:

- `_cons`: For `order` = `faminc` = 0, the average expected birth weight is 112.76 ounces, i.e. $\approx 3.18kg$.
- `faminc`: Keeping `order` constant, a USD 1,000 increase in family income corresponds to an increase of expected birth weight of 0.12 ounces, i.e. $\approx 3.4g$.
- `order`: For the same family income, the next-born child is expected to be approx. 1.14 ounces ($\approx 32.4g$) heavier than his/her sibling born right before him/her.
- All coefficients are found to be highly significant.

## Birth weight: regression 1

- Look at the test for overall relation.

## Birth weight: regression 1

- Look at the test for overall relation.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |   9103.2286         2  4551.6143   Prob > F        =    0.0000
    Residual |  565508.491     1,385  408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.207
```

- Remember [Slides 4, p.10]: *'One useful special case of the F-test is in testing whether all the explanatory variables taken together are significant in the regression model.'*

# Birth weight: regression 1

- Look at the test for overall relation.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |   9103.2286         2  4551.6143   Prob > F        =    0.0000
    Residual |  565508.491     1,385  408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.207
```

- Remember [Slides 4, p.10]: *'One useful special case of the F-test is in testing whether all the explanatory variables taken together are significant in the regression model.'*

  For our multivariate regression

$$\mathtt{bwght}_\mathtt{i} = \beta_1 + \beta_2 \mathtt{faminc}_\mathtt{i} + \beta_3 \mathtt{order}_\mathtt{i} + u_i$$

  this means $H_0 : \beta_2 = \beta_3 = 0$.

# Birth weight: regression 1

- Look at the test for overall relation.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |   9103.2286         2  4551.6143   Prob > F        =    0.0000
    Residual |  565508.491     1,385  408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.207
```

- Remember [Slides 4, p.10]: *'One useful special case of the F-test is in testing whether all the explanatory variables taken together are significant in the regression model.'*

  For our multivariate regression

$$\texttt{bwght}_{\texttt{i}} = \beta_1 + \beta_2 \texttt{faminc}_{\texttt{i}} + \beta_3 \texttt{order}_{\texttt{i}} + u_i$$

  this means $H_0 : \beta_2 = \beta_3 = 0$.

- This test statistic follows an $F_{(q,n-k)}$ distribution under $H_0$ where $q = 2$ (no. of tested parameters) and $(n-k) = 1,385$.

## Birth weight: regression 1

- Look at the test for overall relation.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |  9103.2286         2  4551.6143     Prob > F        =    0.0000
    Residual |  565508.491     1,385  408.30938    R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |  574611.72      1,387  414.283864   Root MSE        =    20.207
```

- Remember [Slides 4, p.10]: *'One useful special case of the F-test is in testing whether all the explanatory variables taken together are significant in the regression model.'*

  For our multivariate regression

$$\texttt{bwght}_{\texttt{i}} = \beta_1 + \beta_2\texttt{faminc}_{\texttt{i}} + \beta_3\texttt{order}_{\texttt{i}} + u_i$$

  this means $H_0 : \beta_2 = \beta_3 = 0$.

- This test statistic follows an $F_{(q,n-k)}$ distribution under $H_0$ where $q = 2$ (no. of tested parameters) and $(n-k) = 1,385$.
- Compare test statistic against critical value (and/or use p-value): We reject $H_0$ here; variables are 'jointly significant'. Let's do the same 'by hand'!

## Birth weight: regression 1

- Interpret the goodness-of-fit of the regression, i.e. the $R^2$.

## Birth weight: regression 1

- Interpret the goodness-of-fit of the regression, i.e. the $R^2$.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |   9103.2286         2  4551.6143   Prob > F        =    0.0000
    Residual |  565508.491     1,385  408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.207
```

- See Slides 4, p.19-21 for more details.

## Birth weight: regression 1

- Interpret the goodness-of-fit of the regression, i.e. the $R^2$.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |  9103.2286         2  4551.6143   Prob > F        =    0.0000
    Residual |  565508.491     1,385  408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |  574611.72      1,387  414.283864  Root MSE        =    20.207
```

- See Slides 4, p.19-21 for more details.
- Main takeaway here: Our model fits the data only very poorly (very low $R^2$).

## Birth weight: regression 1

- Interpret the goodness-of-fit of the regression, i.e. the $R^2$.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |   9103.2286         2   4551.6143   Prob > F        =    0.0000
    Residual |  565508.491     1,385   408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.207
```

- See Slides 4, p.19-21 for more details.
- Main takeaway here: Our model fits the data only very poorly (very low $R^2$).
- Only about 1.58% of the variation of the dependent variable about its mean is explained by regression 1.

# Birth weight: regression 1

- Interpret the goodness-of-fit of the regression, i.e. the $R^2$.

```
      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(2, 1385)      =     11.15
       Model |   9103.2286         2  4551.6143   Prob > F        =    0.0000
    Residual |  565508.491     1,385  408.30938   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0144
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.207
```

- See Slides 4, p.19-21 for more details.
- Main takeaway here: Our model fits the data only very poorly (very low $R^2$).
- Only about 1.58% of the variation of the dependent variable about its mean is explained by regression 1.
- As before, we can calculate the $R^2$ using other parts of the regression output. See how it's done!

## Birth weight: regression 2

- Estimate a regression of birth weight on a constant, family income, birth order **and cigs**.

## Birth weight: regression 2

- Estimate a regression of birth weight on a constant, family income, birth order **and cigs**.

```
. regress bwght faminc order cigs
```

## Birth weight: regression 2

- Estimate a regression of birth weight on a constant, family income, birth order **and cigs**.

```
. regress bwght faminc order cigs

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(3, 1384)      =     16.63
       Model |  19996.5211         3  6665.50703   Prob > F        =    0.0000
    Residual |  554615.199     1,384  400.733525   R-squared       =    0.0348
-------------+----------------------------------   Adj R-squared   =    0.0327
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.018

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      faminc |   .0979201   .0291868     3.35   0.001      .040665    .1551752
       order |   1.616372    .603955     2.68   0.008     .4316058    2.801138
        cigs |  -.4771537    .091518    -5.21   0.000    -.6566827   -.2976247
       _cons |   114.2143     1.4693    77.73   0.000     111.3321    117.0966
```

- Observe how the regression output has changed!

## Birth weight: regression 2

- Estimate a regression of birth weight on a constant, family income, birth order **and cigs**.

```
. regress bwght faminc order cigs

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(3, 1384)      =     16.63
       Model |  19996.5211         3  6665.50703   Prob > F        =    0.0000
    Residual |  554615.199     1,384  400.733525   R-squared       =    0.0348
-------------+----------------------------------   Adj R-squared   =    0.0327
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.018

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      faminc |   .0979201   .0291868     3.35   0.001      .040665    .1551752
       order |   1.616372    .603955     2.68   0.008     .4316058    2.801138
        cigs |  -.4771537    .091518    -5.21   0.000    -.6566827   -.2976247
       _cons |   114.2143     1.4693    77.73   0.000     111.3321    117.0966
```

- Observe how the regression output has changed!
- Conduct a 1-sided test on the `cigs` coefficient using statistical tables!

## Birth weight: regression 2

One- vs two-sided tests?

- Two-sided:
  - Testing for the possibility of the relationship in both directions.

## Birth weight: regression 2

One- vs two-sided tests?

- Two-sided:
  - Testing for the possibility of the relationship in both directions.
  - IOW A two-sided test will test both if the coefficient is significantly bigger **and** significantly smaller than zero.

## Birth weight: regression 2

One- vs two-sided tests?

- Two-sided:
  - Testing for the possibility of the relationship in both directions.
  - IOW A two-sided test will test both if the coefficient is significantly bigger **and** significantly smaller than zero.
  - Given a fixed significance level of, say, $\alpha = 0.05$, a two-sided test allots half of our $\alpha$ to testing the statistical significance in *one* and the other half of $\alpha$ to the other direction.

## Birth weight: regression 2

One- vs two-sided tests?

- Two-sided:
  - Testing for the possibility of the relationship in both directions.
  - IOW A two-sided test will test both if the coefficient is significantly bigger **and** significantly smaller than zero.
  - Given a fixed significance level of, say, $\alpha = 0.05$, a two-sided test allots half of our $\alpha$ to testing the statistical significance in *one* and the other half of $\alpha$ to the other direction.
  - This means `cigs` is considered significantly different from 0 if the test statistic is in the top 2.5% or bottom 2.5% of its probability distribution, resulting in a p-value less than 0.05.

## Birth weight: regression 2

One- vs two-sided tests?

- One-sided:
  - Tests either if the estimated parameter is significantly greater **or** significantly smaller than zero, but not both.

## Birth weight: regression 2

One- vs two-sided tests?

- One-sided:
    - Tests either if the estimated parameter is significantly greater **or** significantly smaller than zero, but not both.
    - IOW it tests for the possibility of the relationship in one direction and completely disregarding the other direction.

## Birth weight: regression 2

One- vs two-sided tests?

- One-sided:
  - Tests either if the estimated parameter is significantly greater **or** significantly smaller than zero, but not both.
  - IOW it tests for the possibility of the relationship in one direction and completely disregarding the other direction.
  - A one-sided test allots **all** of our $\alpha$ to testing the statistical significance in *one* direction only.

## Birth weight: regression 2

One- vs two-sided tests?

- One-sided:
  - Tests either if the estimated parameter is significantly greater **or** significantly smaller than zero, but not both.
  - IOW it tests for the possibility of the relationship in one direction and completely disregarding the other direction.
  - A one-sided test allots **all** of our $\alpha$ to testing the statistical significance in *one* direction only.
  - This means that, depending on the chosen tail, `cigs` is considered significantly lower (higher) than 0 if the test statistic is in the bottom (top) 5% of its probability distribution, resulting in a p-value less than 0.05.

## Birth weight: regression 2

One- vs two-sided tests?

- One-sided:
  - Tests either if the estimated parameter is significantly greater **or** significantly smaller than zero, but not both.
  - IOW it tests for the possibility of the relationship in one direction and completely disregarding the other direction.
  - A one-sided test allots **all** of our $\alpha$ to testing the statistical significance in *one* direction only.
  - This means that, depending on the chosen tail, `cigs` is considered significantly lower (higher) than 0 if the test statistic is in the bottom (top) 5% of its probability distribution, resulting in a p-value less than 0.05.
  - The one-tailed test provides more power to detect an effect in one direction by not testing the effect in the other direction.

## Birth weight: regression 2

- One-sided test in our example

## Birth weight: regression 2

- One-sided test in our example

```
. regress bwght faminc order cigs
```

## Birth weight: regression 2

- One-sided test in our example

```
. regress bwght faminc order cigs

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      faminc |   .0979201   .0291868     3.35   0.001      .040665    .1551752
       order |   1.616372    .603955     2.68   0.008     .4316058    2.801138
        cigs |  -.4771537    .091518    -5.21   0.000    -.6566827   -.2976247
       _cons |   114.2143     1.4693    77.73   0.000     111.3321    117.0966
```

- Above $H_0 : \beta_{cigs} = 0,\ \ H_a : \beta_{cigs} \neq 0$ (two-sided).
- We want to test for the one-sided alternative that $\beta_{cigs} < 0$. Note that the effect has the correct 'predicted sign'.
- To get the p-value, simply divide $p - Val_{cigs}/2$ because the effect is going in the 'predicted' direction. This is P($|-5.21|$).
- We reject $H_0$ in favour of the one-sided alternative indicating that smoking has a significantly negative impact on birth weight.

## Birth weight: regression 3

- Estimate a regression of birth weight on a constant, family income, birth order, cigs and mothereduc.

## Birth weight: regression 3

- Estimate a regression of birth weight on a constant, family income, birth order, cigs and mothereduc.

```
. regress bwght faminc order cigs mothereduc
```

## Birth weight: regression 3

- Estimate a regression of birth weight on a constant, family income, birth order, cigs and mothereduc.

```
. regress bwght faminc order cigs mothereduc

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =     12.49
       Model |  20031.4556         4  5007.8639   Prob > F        =    0.0000
    Residual |  554580.264     1,383  400.998022   R-squared       =    0.0349
-------------+----------------------------------   Adj R-squared   =    0.0321
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.025

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      faminc |   .0937846   .0323843     2.90   0.004     .0302569    .1573122
       order |   1.631028   .6061915     2.69   0.007     .4418743    2.820183
        cigs |  -.4730827   .0925814    -5.11   0.000    -.6546978   -.2914676
  mothereduc |    .076161   .2580337     0.30   0.768    -.4300187    .5823408
       _cons |    113.317   3.376943    33.56   0.000     106.6925    119.9415
------------------------------------------------------------------------------
```

- Interpret results!

## Model specification

From the lecture [slides 4, pp. 14]:

- If an important explanatory variable is **missing** from our model, the estimated coefficients are **biased** (omitted variable bias)!

# Model specification

From the lecture [slides 4, pp. 14]:

- If an important explanatory variable is **missing** from our model, the estimated coefficients are **biased** (omitted variable bias)!
- If an irrelevant explanatory variable is included in our model, the estimated coefficients are **still unbiased** but **inefficient**, i.e. its variance is greater when the irrelevant variable is included.

## Model specification

From the lecture [slides 4, pp. 14]:

- If an important explanatory variable is **missing** from our model, the estimated coefficients are **biased** (omitted variable bias)!
- If an irrelevant explanatory variable is included in our model, the estimated coefficients are **still unbiased** but **inefficient**, i.e. its variance is greater when the irrelevant variable is included.

Which is the best specification for our problem?

## Model specification

From the lecture [slides 4, pp. 14]:

- If an important explanatory variable is **missing** from our model, the estimated coefficients are **biased** (omitted variable bias)!
- If an irrelevant explanatory variable is included in our model, the estimated coefficients are **still unbiased** but **inefficient**, i.e. its variance is greater when the irrelevant variable is included.

Which is the best specification for our problem?

- regression 1: low $R^2$, `cigs` missing; **probably under-sepcified.**

## Model specification

From the lecture [slides 4, pp. 14]:

- If an important explanatory variable is **missing** from our model, the estimated coefficients are **biased** (omitted variable bias)!
- If an irrelevant explanatory variable is included in our model, the estimated coefficients are **still unbiased** but **inefficient**, i.e. its variance is greater when the irrelevant variable is included.

Which is the best specification for our problem?

- regression 1: low $R^2$, `cigs` missing; **probably under-sepcified.**
- regression 3: `mothereduc` probably irrelevant; **probably over-sepcified.**

## Model specification

From the lecture [slides 4, pp. 14]:

- If an important explanatory variable is **missing** from our model, the estimated coefficients are **biased** (omitted variable bias)!
- If an irrelevant explanatory variable is included in our model, the estimated coefficients are **still unbiased** but **inefficient**, i.e. its variance is greater when the irrelevant variable is included.

Which is the best specification for our problem?

- regression 1: low $R^2$, `cigs` missing; **probably under-sepcified.**
- regression 3: `mothereduc` probably irrelevant; **probably over-sepcified.**
- regression 2: note that $adj.\ R^2_{reg.3} < adj.\ R^2_{reg.2}$; **probably best fit.**

## Homo- vs. heteroskedasticity

See lecture [slides 4, pp. 37]:

## Homo- vs. heteroskedasticity

See lecture [slides 4, pp. 37]:

- **Homoskedasticity**: One of the classical assumptions of the linear regression model [Slides 3, p.3].

## Homo- vs. heteroskedasticity

See lecture [slides 4, pp. 37]:

- **Homoskedasticity**: One of the classical assumptions of the linear regression model [Slides 3, p.3].
- **Heteroskedasticity**: Errors have non-constant variances.

## Homo- vs. heteroskedasticity

See lecture [slides 4, pp. 37]:

- **Homoskedasticity**: One of the classical assumptions of the linear regression model [Slides 3, p.3].
- **Heteroskedasticity**: Errors have non-constant variances.
    - Example: More variation in consumption behaviour for rich compared to poor families.

## Homo- vs. heteroskedasticity

See lecture [slides 4, pp. 37]:

- **Homoskedasticity**: One of the classical assumptions of the linear regression model [Slides 3, p.3].
- **Heteroskedasticity**: Errors have non-constant variances.
  - Example: More variation in consumption behaviour for rich compared to poor families.
  - Causes: measurement error, sub-population differences, miss-specification (omitted variables) etc. (many causes possible).

## Homo- vs. heteroskedasticity

See lecture [slides 4, pp. 37]:

- **Homoskedasticity**: One of the classical assumptions of the linear regression model [Slides 3, p.3].
- **Heteroskedasticity**: Errors have non-constant variances.
  - Example: More variation in consumption behaviour for rich compared to poor families.
  - Causes: measurement error, sub-population differences, miss-specification (omitted variables) etc. (many causes possible).
  - Consequences: Coefficient estimates remain unbiased. However, OLS is no longer efficient (smallest variance) **and** estimated standard errors are biased.

## Homo- vs. heteroskedasticity

See lecture [slides 4, pp. 37]:

- **Homoskedasticity**: One of the classical assumptions of the linear regression model [Slides 3, p.3].
- **Heteroskedasticity**: Errors have non-constant variances.
  - Example: More variation in consumption behaviour for rich compared to poor families.
  - Causes: measurement error, sub-population differences, miss-specification (omitted variables) etc. (many causes possible).
  - Consequences: Coefficient estimates remain unbiased. However, OLS is no longer efficient (smallest variance) **and** estimated standard errors are biased.
    $\Rightarrow$ Hypothesis testing is invalid!

## Homo- vs. heteroskedasticity

- Re-run regression 2 and check for heteroskedasticity.

## Homo- vs. heteroskedasticity

- Re-run regression 2 and check for heteroskedasticity.

```
. regress bwght faminc order cigs
```

## Homo- vs. heteroskedasticity

- Re-run regression 2 and check for heteroskedasticity.

```
. regress bwght faminc order cigs
```

- Visual analysis first; e.g. for `faminc`

## Homo- vs. heteroskedasticity

- Re-run regression 2 and check for heteroskedasticity.

```
. regress bwght faminc order cigs
```

- Visual analysis first; e.g. for `faminc`

```
. rvpplot faminc
```

## Homo- vs. heteroskedasticity

- Re-run regression 2 and check for heteroskedasticity.

```
. regress bwght faminc order cigs
```

- Visual analysis first; e.g. for `faminc`

```
. rvpplot faminc
```

- White test for heteroskedasticity:

## Homo- vs. heteroskedasticity

- Re-run regression 2 and check for heteroskedasticity.

```
. regress bwght faminc order cigs
```

- Visual analysis first; e.g. for `faminc`

```
. rvpplot faminc
```

- White test for heteroskedasticity:

```
. estat imtest, white
```

## Homo- vs. heteroskedasticity

- Re-run regression 2 and check for heteroskedasticity.

```
. regress bwght faminc order cigs
```

- Visual analysis first; e.g. for `faminc`

```
. rvpplot faminc
```

- White test for heteroskedasticity:

```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(9)      =      0.63
         Prob > chi2  =    0.9999
```

- We fail to reject $H_0$ of homoskedasticity; no statistical evidence of heteroskedasticity.
- This is in line with our visual analysis.

# Thank you and see you in two weeks!

**Juergen Amann**

**juergen.amann@nottingham.ac.uk**
**Wednesday 12:00 - 13:00, C42 SCGB**

## Overall relation 'by hand' ◂ Go back

There are two ways how we can recover the previous statistic ourselves:

## Overall relation 'by hand' ◂ Go back

There are two ways how we can recover the previous statistic ourselves:

1. Joint parameter test (remember, we did this last week!)

```
. test faminc = order = 0

 ( 1)  faminc - order = 0
 ( 2)  faminc = 0

       F(  2,  1385) =   11.15
            Prob > F =    0.0000
```

# Overall relation 'by hand' Go back

There are two ways how we can recover the previous statistic ourselves:

1. Joint parameter test (remember, we did this last week!)

```
. test faminc = order = 0

 ( 1)  faminc - order = 0
 ( 2)  faminc = 0

       F(  2,  1385) =   11.15
            Prob > F =    0.0000
```

2. Use one of the results from lecture[Slides 4, p.11], e.g.: $F^* = \frac{R^2/(q)}{(1-R^2)/(n-k)}$

## Overall relation 'by hand' Go back

There are two ways how we can recover the previous statistic ourselves:

1. Joint parameter test (remember, we did this last week!)

```
. test faminc = order = 0

 ( 1)  faminc - order = 0
 ( 2)  faminc = 0

       F(  2,  1385) =   11.15
            Prob > F =    0.0000
```

2. Use one of the results from lecture[Slides 4, p.11], e.g.: $F^* = \frac{R^2/(q)}{(1-R^2)/(n-k)}$

```
. display 0.0158/(1 - 0.0158) * 1385/2
11.117151
```

## $R^2$ 'by hand' ◂ Go back

From the lecture slides [Slides 3, pp.20]:

- The $R^2$ measure of goodness-of-fit is defined by $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$.

## $R^2$ 'by hand' Go back

From the lecture slides [Slides 3, pp.20]:

- The $R^2$ measure of goodness-of-fit is defined by $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$.

```
. regress bwght faminc order cigs mothereduc
```

## $R^2$ 'by hand' Go back

From the lecture slides [Slides 3, pp.20]:

- The $R^2$ measure of goodness-of-fit is defined by $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$.

```
. regress bwght faminc order cigs mothereduc

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =     12.49
       Model |  20031.4556         4  5007.8639   Prob > F        =    0.0000
    Residual |  554580.264     1,383  400.998022   R-squared       =    0.0349
-------------+----------------------------------   Adj R-squared   =    0.0321
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.025
```

## $R^2$ 'by hand' Go back

From the lecture slides [Slides 3, pp.20]:

- The $R^2$ measure of goodness-of-fit is defined by $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$.

```
. regress bwght faminc order cigs mothereduc

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =     12.49
       Model |  20031.4556         4  5007.8639   Prob > F        =    0.0000
    Residual |  554580.264     1,383  400.998022   R-squared       =    0.0349
-------------+----------------------------------   Adj R-squared   =    0.0321
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.025

. display 20031.4556 /  574611.72
0.03486085
```

## $R^2$ 'by hand' Go back

From the lecture slides [Slides 3, pp.20]:

- The $R^2$ measure of goodness-of-fit is defined by $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$.

```
. regress bwght faminc order cigs mothereduc

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =     12.49
       Model |  20031.4556         4  5007.8639   Prob > F        =    0.0000
    Residual |  554580.264     1,383  400.998022   R-squared       =    0.0349
-------------+----------------------------------   Adj R-squared   =    0.0321
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.025

. display 20031.4556 /  574611.72
0.03486085
```

- The adjusted $R^2$ incorporates a penalty for including additional variables: $adj.R^2 = 1 - \left[\left(\frac{T-1}{T-k}\right)(1 - R^2)\right]$

## $R^2$ 'by hand' Go back

From the lecture slides [Slides 3, pp.20]:

- The $R^2$ measure of goodness-of-fit is defined by $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$.

```
. regress bwght faminc order cigs mothereduc

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(4, 1383)      =     12.49
       Model |  20031.4556         4  5007.8639   Prob > F        =    0.0000
    Residual |  554580.264     1,383  400.998022   R-squared       =    0.0349
-------------+----------------------------------   Adj R-squared   =    0.0321
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.025

. display 20031.4556 /  574611.72
0.03486085
```

- The adjusted $R^2$ incorporates a penalty for including additional variables: $adj.R^2 = 1 - \left[\left(\frac{T-1}{T-k}\right)(1 - R^2)\right]$

```
. display 1 - ((1388-1)/(1383)*(1-20031.4556/574611.72))
0.03206942
```