Empirical Banking and Finance
Tutorial 3 Solution

Konrad Adler

Institute for Financial Economics and Statistics
University of Bonn

Summer 2022

- Mostly satisfied with the solutions submitted
- - solution with unacceptable formatting
- - some left out the first questions
- In view of the exam, do answer all questions asked.
- When I ask about the mechanism I expect a description that X causes Y etc.

This Lecture

- Solution
- Combining Instruments

1. Motivating the use of Instrumental Variables (IV)
    a) In a cross-country regression with average *gdpgrowth* from 1960-1995 on the LHS and *private_credit_1960* on the RHS mention two distinct economic mechanisms why the estimated coefficient might *not* represent the causal effect of "finance" on subsequent gdp growth.
        - # 1 Anticipation:

1. Motivating the use of Instrumental Variables (IV)
    a) In a cross-country regression with average *gdpgrowth* from 1960-1995 on the LHS and *private_credit_1960* on the RHS mention two distinct economic mechanisms why the estimated coefficient might *not* represent the causal effect of "finance" on subsequent gdp growth.
        - # 1 Anticipation:
        - Levine (2005) writes that "While King and Levine (1993) and Levine and Zervos (1998) show that financial development predicts economic growth, these results do not settle the issue of causality. It may simply be the case that financial markets develop in anticipation of future economic activity. Thus, finance may be a leading indicator rather than a fundamental cause."

1. Motivating the use of Instrumental Variables (IV)
    a) In a cross-country regression with average *gdpgrowth* from 1960-1995 on the LHS and *private_credit_1960* on the RHS mention two distinct economic mechanisms why the estimated coefficient might *not* represent the causal effect of "finance" on subsequent gdp growth.
        - # 1 Anticipation:
        - Levine (2005) writes that "While King and Levine (1993) and Levine and Zervos (1998) show that financial development predicts economic growth, these results do not settle the issue of causality. It may simply be the case that financial markets develop in anticipation of future economic activity. Thus, finance may be a leading indicator rather than a fundamental cause."

        - As the quote above states, financial markets are forward looking.
        - Credit to GDP might increase in anticipation of future growth
        - If these anticipations are correct on average, both future *gdpgrowth* and current *private_credit_1960* are driven by, for example, increases in productivity
        - Example: Oil reserves found

1. Motivating the use of Instrumental Variables (IV)
    a) In a cross-country regression with average *gdpgrowth* from 1960-1995 on the LHS and *private_credit_1960* on the RHS mention two distinct economic mechanisms why the estimated coefficient might *not* represent the causal effect of "finance" on subsequent gdp growth.
        - # 2 Institutions:
        - The quality of institutions of a country might at the same time lead to a high initial credit/GDP and also increase subsequent GDP growth
        - Omitted variable at the country level
        - Nice idea by one of you: quality of credit

1. Motivating the use of Instrumental Variables (IV)

   b) Take one of the two mechanisms from a) and use the omitted variable bias formula to determine the direction of the bias introduced. Explain your reasoning.

$$gdpgrowth_{t+1} = \beta_0 + \beta_1 private\_credit\_1960_t + \beta_2 \text{good institution}_t$$

$$gdpgrowth_{t+1} = \tilde{\beta}_0 + \tilde{\beta}_1 private\_credit\_1960_t$$

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \delta_{private\_credit\_1960, \text{good institution}}$$

$$\text{good institution}_t = \delta_0 + \delta_{private\_credit\_1960, \text{good institution}} private\_credit\_1960_t$$

- Good institutions are likely to increase GDP growth: $\beta_2 > 0$
- Good institutions are likely associated with higher credit/GDP: $\delta_{private\_credit\_1960, \text{good institution}} > 0$
- Therefore, $\tilde{\beta}_1 > \beta_1$, we overestimate the effect of *private_credit_1960* on $gdpgrowth_{t+1}$ when we do not include $\text{good institution}_t$ in the regression

1. Motivating the use of Instrumental Variables (IV)

   c) How could an instrumental variable approach solve the issue of causality?

   - The main idea of an IV approach is to replace the endogenous RHS variable, in our case *private_credit_1960*, with an instrument z, which is strongly correlated with the endogenous RHS variable...
   - ...but, and this is where z differs from the endogenous variable, the instrument should not be correlated with any other determinant of the outcome variable
   - An IV would help to address this reverse causality problem by only considering the impact of financial development on growth that is "exogenous".

$$\begin{aligned} cov(y_i, z_i) &= cov(\beta_0 + \beta_1 R_i + \beta_2 x_i + e_i, z) \\ &= 0 + \beta_1 \underbrace{cov(R_i, z_i)}_{\substack{>0 \\ \text{correlated with} \\ \text{variable of interest}}} + \underbrace{cov(\beta_2 x_i + e_i, z_i)}_{\substack{=0 \\ \text{exclusion} \\ \text{restriction}}} \end{aligned}$$

$$\beta_1 = \underbrace{\frac{cov(y_i, z_i)}{cov(R_i, z_i)}}_{\text{We know those}}$$

2. Regression 1: OLS

   a) Run a regression of *gdpgrowth* on *private_credit_1960*

```
. * a)
. reg gdpgrowth private_credit_1960 , robust

Linear regression                               Number of obs     =         83
                                                F(1, 81)          =      10.80
                                                Prob > F          =     0.0015
                                                R-squared         =     0.1059
                                                Root MSE          =     .02369

------------------------------------------------------------------------------
             |               Robust
  gdpgrowth  |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
private_cre~1960 |   .0354527   .0107892     3.29   0.002     .0139856    .0569199
       _cons |   .0076595   .0041493     1.85   0.069    -.0005962    .0159152
------------------------------------------------------------------------------
```

2. Regression 1: OLS

   b) Very briefly comment on the coefficient of *private_credit_1960*: Whether it is significant, its size and sign. Provide a 95% confidence interval for the coefficient.

      - The coefficient is positive and statistically significant at the 1% level.
      - A 1 sd increase in *private_credit_1960* increases subsequent GDP growth by 0.325 sd.
      - With a 95% chance the true coefficient lies between 0.014 and 0.057

3. The Instruments

   We now use the national legal origin (Porta et al. (1998)) of a country as an instrument for its financial development

   a) Discuss whether a country's legal origin might or not satisfy the exclusion restriction? Provide two arguments/mechanisms in favor, and two arguing against the assumption.

      - Pro:

3. The Instruments

   We now use the national legal origin (Porta et al. (1998)) of a country as an instrument for its financial development

   a) Discuss whether a country's legal origin might or not satisfy the exclusion restriction? Provide two arguments/mechanisms in favor, and two arguing against the assumption.

      - Pro:
        1) In the case of colonies, the legal system was forced on the country "randomly", not taking into account local conditions which might affect GDP growth
        2) Legal system might affect the *level* of GDP through other channels than finance, but not GDP growth

      - Contra:

3. The Instruments

   We now use the national legal origin (Porta et al. (1998)) of a country as an instrument for its financial development

   a) Discuss whether a country's legal origin might or not satisfy the exclusion restriction? Provide two arguments/mechanisms in favor, and two arguing against the assumption.

      - Pro:
        1) In the case of colonies, the legal system was forced on the country "randomly", not taking into account local conditions which might affect GDP growth
        2) Legal system might affect the *level* of GDP through other channels than finance, but not GDP growth

      - Contra:
      - When a country was forced to adopt/adopted a legal system it probably also changed other institutions/issues as well which might affect GDP growth through another channel than finance

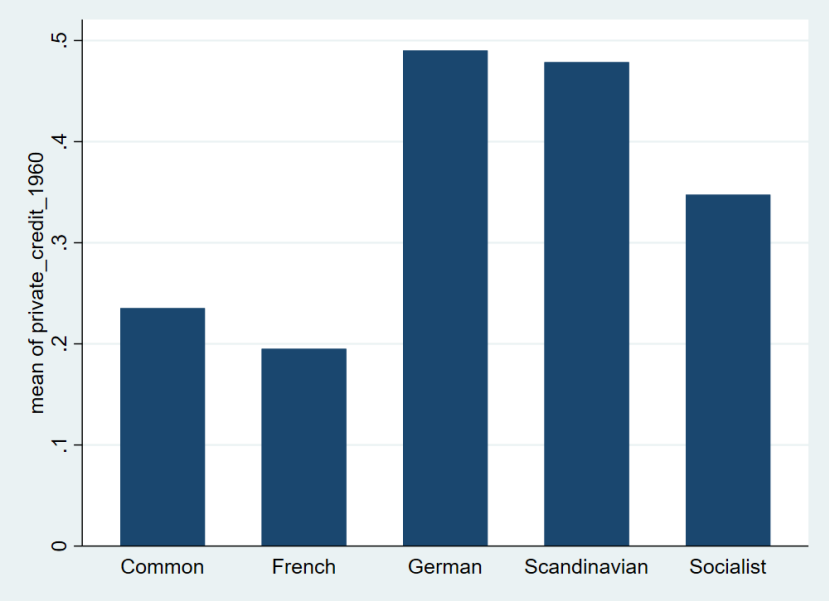
        1) Former french colonies seem to have inherited worse infrastructure relative to other former colonies. Infrastructure quality affects gdpgrowth directly.
        2) Legal origin might also affect patenting law. Patenting affects gdpgrowth directly through its impact on technological progress and not (only) via external finance.

      - Other ideas: Tax code, war, colonizer (extractive vs population colonies), Foreign direct investment, many solutions mentioned schooling...no super convinced by that

3. The Instrument

We now use the national legal origin (Porta et al. (1998)) of a country as an instrument for its financial development

b) Compute average *private_credit_1960* for each group of legal origin. There are several ways to do this. The variable *legor* might be useful. Comment your results.



c) Can we learn anything from the answer to the previous question b) about whether legal origin is a good or a bad instrument?

- Potentially relevant instrument: some variation across the five categories: This indicates that the first stage might work, i.e. legal system can explain variation in credit to GDP.
- Not great: countries of French and UK legal origin have similar levels of credit/GDP

4. Regression 2: IV with one instrument

   a) Use French legal origin *legor_fr* as an instrument for *private_credit_1960*. Compute an IV estimate of the impact of *private_credit_1960* on *gdpgrowth* using the Wald estimator. Explain what your are doing at each step of the calculation.

$$\beta_1 = \frac{\mathbb{E}[gdpgrowth_i | z_i = 1] - \mathbb{E}[gdpgrowth_i | z_i = 0]}{\mathbb{E}[private\_credit\_1960_i | z_i = 1] - \mathbb{E}[private\_credit\_1960_i | z_i = 0]}$$

4. Regression 2: IV with one instrument

   a) Use French legal origin *legor_fr* as an instrument for *private_credit_1960*. Compute an IV estimate of the impact of *private_credit_1960* on *gdpgrowth* using the Wald estimator. Explain what your are doing at each step of the calculation.

$$\beta_1 = \frac{\mathbb{E}[gdpgrowth_i|z_i = 1] - \mathbb{E}[gdpgrowth_i|z_i = 0]}{\mathbb{E}[private\_credit\_1960_i|z_i = 1] - \mathbb{E}[private\_credit\_1960_i|z_i = 0]}$$

- **Numerator:** compare *gdpgrowth* of countries with French legal origin against *gdpgrowth* of countries with other legal origins (we don't take into account about credit/GDP)
  $= -0.015$ countries with French legal origin have 1.5 percentage points lower GDP growth on average
- the overall mean of *gdpgrowth* is 1.7% - the difference is enormous!

4. Regression 2: IV with one instrument

   a) Use French legal origin *legor_fr* as an instrument for *private_credit_1960*. Compute an IV estimate of the impact of *private_credit_1960* on *gdpgrowth* using the Wald estimator. Explain what your are doing at each step of the calculation.

$$\beta_1 = \frac{\mathbb{E}[gdpgrowth_i|z_i = 1] - \mathbb{E}[gdpgrowth_i|z_i = 0]}{\mathbb{E}[private\_credit\_1960_i|z_i = 1] - \mathbb{E}[private\_credit\_1960_i|z_i = 0]}$$

- **Numerator:** compare *gdpgrowth* of countries with French legal origin against *gdpgrowth* of countries with other legal origins (we don't take into account about credit/GDP)
  $= -0.015$ countries with French legal origin have 1.5 percentage points lower GDP growth on average
- the overall mean of *gdpgrowth* is 1.7% - the difference is enormous!

- **Denominator:** Adjust for the difference in terms of *private_credit_1960*
  $= -0.120$ countries with French legal origin have 12 percentage points lower *private_credit_1960* on average
- the overall mean of *private_credit_1960* is 25%
- $\beta_{1,Wald} = 0.122654$

4. Regression 2: IV with one instrument

   a) Use French legal origin *legor_fr* as an instrument for *private_credit_1960*. Compute an IV estimate of the impact of *private_credit_1960* on *gdpgrowth* using the Wald estimator. Explain what your are doing at each step of the calculation.

$$\beta_1 = \frac{\mathbb{E}[gdpgrowth_i | z_i = 1] - \mathbb{E}[gdpgrowth_i | z_i = 0]}{\mathbb{E}[private\_credit\_1960_i | z_i = 1] - \mathbb{E}[private\_credit\_1960_i | z_i = 0]}$$

   - Reminder:
   - What is the problem when the denominator is 0?
   - What is the problem when the difference in the denominator is enormous?

4. Regression 2: IV with one instrument

   a) Use French legal origin *legor_fr* as an instrument for *private_credit_1960*. Compute an IV estimate of the impact of *private_credit_1960* on *gdpgrowth* using the Wald estimator. Explain what your are doing at each step of the calculation.

$$\beta_1 = \frac{\mathbb{E}[gdpgrowth_i | z_i = 1] - \mathbb{E}[gdpgrowth_i | z_i = 0]}{\mathbb{E}[private\_credit\_1960_i | z_i = 1] - \mathbb{E}[private\_credit\_1960_i | z_i = 0]}$$

   - Reminder:
   - What is the problem when the denominator is 0?
   - What is the problem when the difference in the denominator is enormous?
   - When the difference is "too large" it becomes harder to argue that the two groups are comparable.

4. Regression 2: IV with one instrument

   b) Run a Two-Stage-Least-Squares (2SLS) version of Regression 1 using French legal origin *legor_fr* as an instrument for *private_credit_1960*.

```
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                Number of obs =        83
                                                F(  1,    81) =      4.58
                                                Prob > F      =    0.0353
Total (centered) SS     =  .0508324241          Centered R2   =   -0.5348
Total (uncentered) SS   =  .0737812889          Uncentered R2 =   -0.0574
Residual SS             =  .0780155595          Root MSE      =    .03066

------------------------------------------------------------------------------
             |               Robust
   gdpgrowth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
private_cre~1960 |    .122654   .0565947     2.17   0.030     .0117304    .2335775
       _cons |   -.0143999   .0138792    -1.04   0.299    -.0416027    .0128029
------------------------------------------------------------------------------
```

   c) Compare the Wald estimate to the 2SLS estimate.

      - They are the same!

4. Regression 2: IV with one instrument

   d) Compare the coefficient on *private_credit_1960* to the one in the OLS regression. How does the difference between IV and OLS coefficient compare to your answer in question 1 b)?

```
. esttab reg1 reg2

------------------------------------------------------------
                      (1)             (2)
                gdpgrowth       gdpgrowth
------------------------------------------------------------
private~1960       0.0355**          0.123*
                    (3.29)          (2.17)

_cons              0.00766         -0.0144
                    (1.85)         (-1.04)
------------------------------------------------------------
N                       83              83
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

   - Sign: the same
   - Significance: the IV coefficient is less statistically significant than the OLS coefficient, surprising?

4. Regression 2: IV with one instrument

   d) Compare the coefficient on *private_credit_1960* to the one in the OLS regression. How does the difference between IV and OLS coefficient compare to your answer in question 1 b)?

```
. esttab reg1 reg2

------------------------------------------------------------
                      (1)             (2)
                gdpgrowth       gdpgrowth
------------------------------------------------------------
private~1960       0.0355**          0.123*
                   (3.29)          (2.17)

_cons             0.00766         -0.0144
                   (1.85)         (-1.04)
------------------------------------------------------------
N                       83              83
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

- Sign: the same
- Significance: the IV coefficient is less statistically significant than the OLS coefficient, surprising? No, even with strong instruments (to be checked below) 2SLS is less efficient than OLS
- Size: the IV coefficient is larger, suprising?

4. Regression 2: IV with one instrument

d) Compare the coefficient on *private_credit_1960* to the one in the OLS regression. How does the difference between IV and OLS coefficient compare to your answer in question 1 b)?

```
. esttab reg1 reg2

------------------------------------------------------------
                      (1)             (2)
                gdpgrowth       gdpgrowth
------------------------------------------------------------
private~1960       0.0355**          0.123*
                    (3.29)          (2.17)

_cons              0.00766         -0.0144
                    (1.85)         (-1.04)
------------------------------------------------------------
N                       83              83
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

- Sign: the same
- Significance: the IV coefficient is less statistically significant than the OLS coefficient, surprising? No, even with strong instruments (to be checked below) 2SLS is less efficient than OLS
- Size: the IV coefficient is larger, suprising? Yes, because the omitted variable bias formula in 1b) told us the OLS coefficient is biased upwards.

4. Regression 2: IV with one instrument
    e) Provide a 95% confidence interval for the coefficient.
        - $[0.012, 0.234]$
    f) Is the model underidentified, exactly identified or overidentified?
        - Exactly identified

4. Regression 2: IV with one instrument

g) Test whether *legor_fr* is a valid instrument. Provide H_0, H_A, the test statistic, its distribution and the result of the test.

- **First Requirement:** we can look at the p-value of the first stage ("informal test")

```
. reg private_credit_1960  legor_fr , robust

Linear regression                               Number of obs     =         83
                                                F(1, 81)          =       5.78
                                                Prob > F          =     0.0185
                                                R-squared         =     0.0698
                                                Root MSE          =    .22178

------------------------------------------------------------------------------
             |               Robust
private~1960 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    legor_fr |  -.1200627   .0499485    -2.40   0.019    -.2194445   -.0206809
       _cons |   .3151728   .0461621     6.83   0.000     .2233248    .4070209
------------------------------------------------------------------------------
```

- H_0 : *legor_fr* $= 0$
- H_A : *legor_fr* $\neq 0$
- T-stat: -2.40
- Student t distribution
- We reject H_0 at 5%

4. Regression 2: IV with one instrument

g) Test whether *legor_fr* is a valid instrument. Provide H_0, H_A, the test statistic, its distribution and the result of the test.

- **First Requirement:** we can look at the p-value of the first stage ("informal test")

```
. reg private_credit_1960  legor_fr , robust

Linear regression                               Number of obs     =         83
                                                F(1, 81)          =       5.78
                                                Prob > F          =     0.0185
                                                R-squared         =     0.0698
                                                Root MSE          =    .22178

------------------------------------------------------------------------------
             |               Robust
private~1960 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    legor_fr |  -.1200627   .0499485    -2.40   0.019    -.2194445   -.0206809
       _cons |   .3151728   .0461621     6.83   0.000     .2233248    .4070209
------------------------------------------------------------------------------
```

- H_0 : *legor_fr* $= 0$
- H_A : *legor_fr* $\neq 0$
- T-stat: -2.40
- Student t distribution
- We reject H_0 at 5%
- But, this is less significant than what the rule of thumb suggests (p-value < 0.0016) $\rightarrow$ seems that we have a weak instrument

4. Regression 2: IV with one instrument

g) Test whether *legor_fr* is a valid instrument. Provide H_0, H_A, the test statistic, its distribution and the result of the test. (continued)

- **First Requirement:** Stock-Yogo critical values ("formal test")

```
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                Number of obs =        83
                                                F(  1,    81) =      4.58
                                                Prob > F      =    0.0353
Total (centered) SS     =  .0508324241          Centered R2   =   -0.5348
Total (uncentered) SS   =  .0737812889          Uncentered R2 =   -0.0574
Residual SS             =  .0780155595          Root MSE      =    .03066

------------------------------------------------------------------------------
             |               Robust
   gdpgrowth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
private_credit_1960 |    .122654   .0565947     2.17   0.030     .0117304    .2335775
       _cons |   -.0143999   .0138792    -1.04   0.299    -.0416027    .0128029
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):             5.524
                                                  Chi-sq(1) P-val =    0.0188
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                6.073
                         (Kleibergen-Paap rk Wald F statistic):          5.778
Stock-Yogo weak ID test critical values: 10% maximal IV size            16.38
                                         15% maximal IV size             8.96
                                         20% maximal IV size             6.66
                                         25% maximal IV size             5.53
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
```

4. Regression 2: IV with one instrument

   g) Test whether *legor_fr* is a valid instrument. Provide H_0, H_A, the test statistic, its distribution and the result of the test. (continued)

      - **First Requirement:** Stock-Yogo critical values ("formal test")
      - The first stage F-stats are 6.1 and 5.8
      - "Regular" critical value is $F_{1,82} \approx 4$ with $\alpha = 0.05$
      - Stock - Yogo critical values are 5.5 (25% maximal IV size) and 6.7 (20% maximal IV size)
        $\rightarrow$ Given our (relatively weak) first stage we will have to accept an actual size between 20 and 25% when testing the second stage coefficient of interest

      - **Second Requirement:** Exclusion restriction?

4. Regression 2: IV with one instrument

   g) Test whether *legor_fr* is a valid instrument. Provide H_0, H_A, the test statistic, its distribution and the result of the test. (continued)

      - **First Requirement:** Stock-Yogo critical values ("formal test")
      - The first stage F-stats are 6.1 and 5.8
      - "Regular" critical value is $F_{1,82} \approx 4$ with $\alpha = 0.05$
      - Stock - Yogo critical values are 5.5 (25% maximal IV size) and 6.7 (20% maximal IV size)
        $\rightarrow$ Given our (relatively weak) first stage we will have to accept an actual size between 20 and 25% when testing the second stage coefficient of interest

      - **Second Requirement:** Exclusion restriction? - can only be tested formally when we have more than 1 instrument

5. Regression 3: IV with several instruments

   a) Run an 2SLS version of Regression 1 using four out of five legal origin dummies as instruments for *private_credit_1960*

```
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                Number of obs =        83
                                                F(  1,    81) =      4.88
                                                Prob > F      =    0.0300
Total (centered) SS     =   .0508324241         Centered R2   =   -0.0636
Total (uncentered) SS   =   .0737812889         Uncentered R2 =    0.2672
Residual SS             =   .0540666019         Root MSE      =    .02552

------------------------------------------------------------------------------
             |               Robust
   gdpgrowth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
private_cred~1960 |   .0803088   .0359093     2.24   0.025     .0099278    .1506898
       _cons |  -.0036878   .0090462    -0.41   0.684    -.0214181    .0140425
------------------------------------------------------------------------------
```

5. Regression 3: IV with several instruments

   b) Why cannot all legal origin dummies be included?
      - Multicollinearity problem in the first stage

   c) Compare the coefficient on *private_credit_1960* to the one in question 4 b). Provid a brief comment.

```
. esttab reg1 reg2 reg3

------------------------------------------------------------
                      (1)             (2)             (3)
                gdpgrowth       gdpgrowth       gdpgrowth
------------------------------------------------------------
private~1960       0.0355**          0.123*          0.0803*
                    (3.29)          (2.17)          (2.24)

_cons              0.00766         -0.0144        -0.00369
                    (1.85)         (-1.04)         (-0.41)
------------------------------------------------------------
N                       83              83              83
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

   - Size: even with all the instruments the IV coefficient is larger than the OLS coefficient, the direction of the bias found in 1 b) is not confirmed
   - Sign: the same
   - Significance: Both IV coefficients are less significant than the OLS coefficient

5. Regression 3: IV with several instruments

   d) Provide a 95% confidence interval for the coefficient.

      - $[0.01, 0.15]$ (smaller than the one found using only one instrument $[0.012, 0.234]$)

   e) Is the model underidentified, exactly identified or overidentified?

5. Regression 3: IV with several instruments

   d) Provide a 95% confidence interval for the coefficient.
      - $[0.01, 0.15]$ (smaller than the one found using only one instrument $[0.012, 0.234]$)

   e) Is the model underidentified, exactly identified or overidentified?
      - Overidentified

   f) Formally test whether the instruments are valid. Provide H_0, H_A, the test statistic, its distribution and the result of the test.

5. Regression 3: IV with several instruments

   d) Provide a 95% confidence interval for the coefficient.
      - $[0.01, 0.15]$ (smaller than the one found using only one instrument $[0.012, 0.234]$)

   e) Is the model underidentified, exactly identified or overidentified?
      - Overidentified

   f) Formally test whether the instruments are valid. Provide H_0, H_A, the test statistic, its distribution and the result of the test.
      - **First Requirement:**
      - F-test of whether the four instruments are jointly significant in the first stage.
      - H_0: $\beta_{legalorigin1} = 0$ and $\beta_{legalorigin2} = 0$ and $\beta_{legalorigin3} = 0$ and $\beta_{legalorigin3} = 0$
      - H_A: H_0 not true, any or all of the coefficients are different from zero
      - F-Stat is 9.89, H_0 of all of the instruments being zero is rejected at 5%, but just below the rule of thumb of an F-stat of 10
      - Still weak instruments

5. Regression 3: IV with several instruments

   g) Formally test whether the instruments are valid. Provide H_0, H_A, the test statistic, its distribution and the result of the test. (continued)

   - **Second Requirement:**
   - Overidentification test of all instruments

```
Hansen J statistic (overidentification test of all instruments):        3.002
                                                  Chi-sq(3) P-val =     0.3913
```

   - H_0 all instruments are exogenous
   - H_A at least one instrument is not exogenous
   - $\chi^2(3) = 3.002$
   - We cannot reject H_0 here, therefore all instruments are exogenous
   - At least formally, never claim that your instruments are exogenous only based on this test

5. Regression 3: IV with several instruments

    h) Is the formal test for the exogeneity of instruments useful in this setting?

5. Regression 3: IV with several instruments

   h) Is the formal test for the exogeneity of instruments useful in this setting?
      - Not useful: the test requires that one instrument is truly exogenous, but with the legal origin dummies either all or none are truly endogenous

   i) Test whether *private_credit_1960* is endogenous. Provide H_0, H_A, the test statistic, its distribution and the result of the test. Why does this matter from an econometric point of view? In your answer, refer to results in previous question(s).
      - H_0: *private_credit_1960* is exogenous.
      - H_A: *private_credit_1960* is endogenous.
      - F-stat of 4.20
      - $\chi^2(1) = 2.99$
      - F-test: we can reject H_0 and conclude that *private_credit_1960* is endogenous at 5%
      - $\chi^2(1)$ test: we cannot reject H_0 and conclude that *private_credit_1960* could be exogenous at 5% .
      - This matters because IV is less efficient than OLS and we should only run IV when it is really necessary.

6. Regression 4: IV with several instruments and several endogenous variables
    a) Run an IV version of Regression 1 using four out of five legal origin dummies as instruments for *private_credit_1960*, but now, add *public_banks_1970* as an additional endogenous dependent variable.
        -

This Lecture

Solution

Combining Instruments

Legal origin dummies

Combining several instruments

- This short section tries to give some intuition how the information of several instruments is aggregated
- With a dummy IV we have seen that 2SLS = Wald estimator
- How are Wald estimators related to 2SLS with several dummy IVs?

Reference: Angrist (2009) 4.1.3

several dummy IV: 2SLS

First Stage

$$\begin{bmatrix} private\_credit\_1960_1 \\ \vdots \\ private\_credit\_1960_N \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & x_1 \\ 1 & 1 & 0 & 0 & x_2 \\ 1 & 0 & 1 & 0 & x_3 \\ 1 & 0 & 1 & 0 & x_4 \\ 1 & 0 & 0 & 1 & x_5 \\ 1 & 0 & 0 & 1 & x_6 \\ 1 & \underbrace{0}_{\text{German}} & \underbrace{0}_{\text{French}} & \underbrace{0}_{\text{Common}} & x_7 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{German} \\ \beta_{French} \\ \beta_{Common} \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$$

several dummy IV: 2SLS

Second Stage

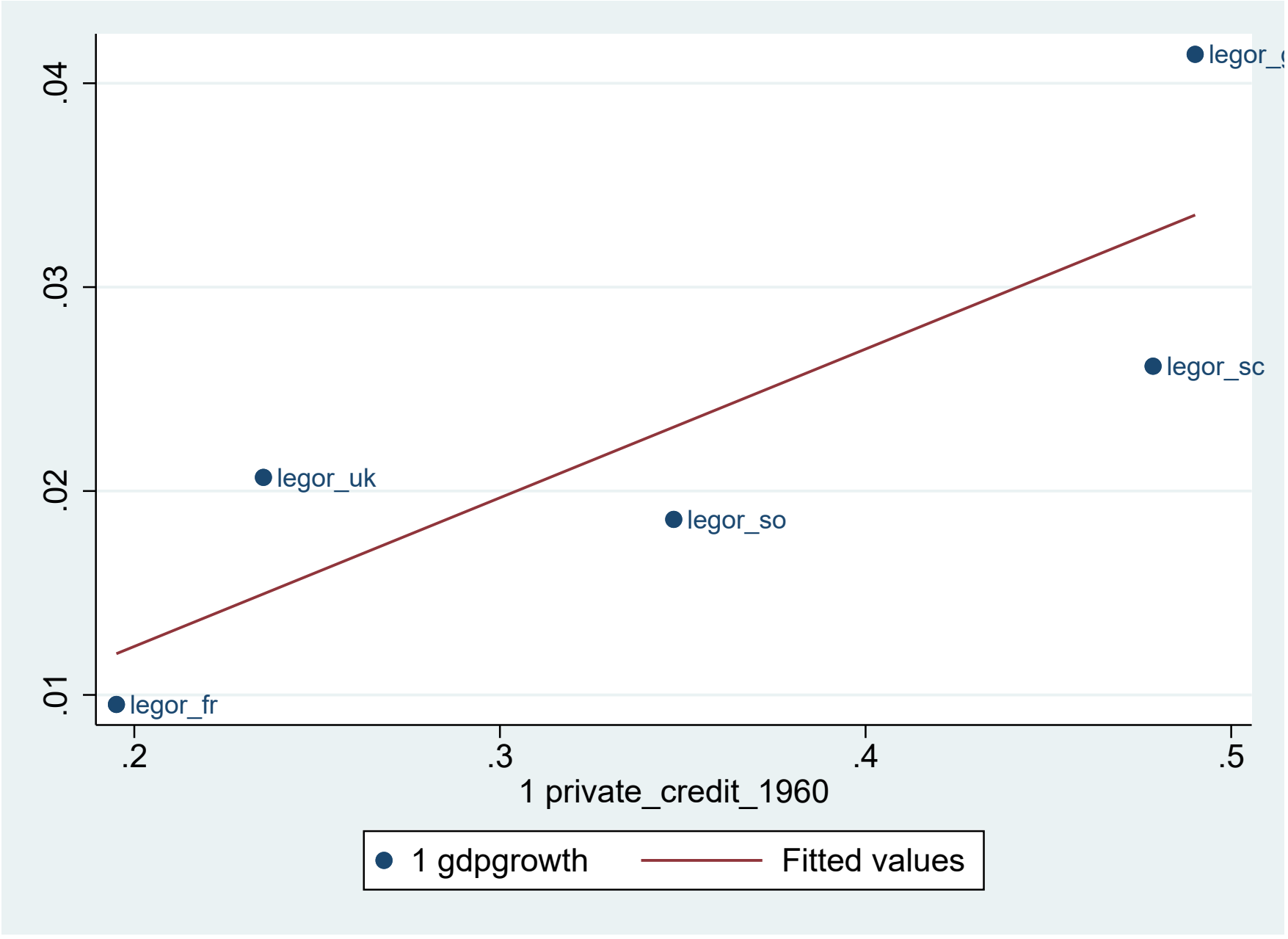
$$y_i = \beta_0 + .0803088 \widehat{private\_credit\_1960}_i + e_i$$

- Combine Wald estimators?
- Wald estimator for each instrument separately?

several dummy IV: Combine Wald estimators

- The 2SLS can be constructed by aggregating IV estimates
- IV for each legal origin category separately
- How does this work?
  i) Compute $\mathbb{E}[y_i|z_i = 1]$
  ii) Compute $\mathbb{E}[FinDev_i|z_i = 1]$
  iii) Do this for all five z_i
  iv) Run a regression through the computed averages of gdp growth and financial development

several dummy IV: Combine Wald estimators



several dummy IV: Combine Wald estimators

- The slope, after reweighting, gives exactly the 2SLS estimate
- Implications
  - We don't need "micro" data to estimate IV
  - Group averages by instrument are enough
- How well does the line fit?
  - Test of whether the parameter of interest should be constant or not
- Take-away
  - 2SLS with several dummy instruments is just a linear combination of each separate IV estimate

several dummy IV: Separate Wald estimators

Wald estimator

$$\beta_1 = \frac{\mathbb{E}[y_i|z_i = 1] - \mathbb{E}[y_i|z_i = 0]}{\mathbb{E}[\mathrm{R}_i|z_i = 1] - \mathbb{E}[\mathrm{R}_i|z_i = 0]}$$

- The assumptions about the instruments should hold for each of them
- The results using individual instruments should be similar to the aggregated one

several dummy IV: Separate Wald estimators

legal origin	wald estimator
legor_fr	.12
legor_uk	-.23
legor_so	.02
legor_ge	.10
legor_sc	.04
2SLS	0.08

- The assumptions about the instruments should hold for each of them
- The results using individual instruments should be similar to the aggregated one

Combining instruments

Take-aways

- 1 dummy IV, no covariates
  - 2SLS = Wald estimator
- several dummy IVs, no covariates
  - 2SLS = linear combination of Wald type estimators
- Practical implications
  - IV can be used having only aggregate data
  - Assess the "fit" across instruments

King, R. G. and Levine, R. (1993). Finance and Growth: Schumpeter Might Be Right. *The Quarterly Journal of Economics*, 108(3):717–737.

Levine, R. (2005). Finance and Growth: Theory and Evidence. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 1 of *Handbook of Economic Growth*, chapter 12, pages 865–934. Elsevier.

Porta, R. L., de Silanes, F. L., Shleifer, A., and Vishny, R. W. (1998). Law and Finance. *Journal of Political Economy*, 106(6):1113–1155.