Økonometri A
Efterår 2025

Mandatory Assignment 1

Tax Revenues and Municipal Tax Rates in Denmark

This assignment is the first of three mandatory assignments in Econometrics A. All three assignments must be passed in order to go to the final exam.

The assignment may be answered in groups of max 3 students.

Practical instructions to the assignment
Read the entire assignment before you begin to respond, and answer all the questions.

The answer to the assignment must be presented in a comprehensive report with relevant tables and figures. The front page of the report must be based on the template that is available on Absalon. Fill in the names and study ID of all group members on the front page.

Prepare one Jupyter Notebook-file that generates all tables and figures that appear in your report. The Python-program should produce tables and figures in the same order as they appear in the report. Comments should clearly indicate which table or figure appearing in the report is being produced. Make sure that the ipynb-file can be executed without any errors. The ipynb-file must include the names and study ID of all group members.

To pass a mandatory assignment, it is required that:

- An adequate response is given to all questions in the assignment.
- More than half of the questions are answered correctly.
- Answers are written in a precise and easy to understand language.

The report must not exceed 8 A4 pages with font size 12, line space 1.5 and page margins of 2.5 cm. This includes the main text, tables and figures in the report, but not the front page.

Uploading your report
Each group must hand-in only one report in total. You must upload two files:

1. The report itself must be uploaded as a PDF file.
2. The Jupyter Notebook ipynb-file must be uploaded as well.

If needed, a free PDF converter can be found here: www.pdf995.com

If group members are assigned to different class teachers, upload the report to the class teacher of your choice. Do not upload the same report to more than ones.

Access to data
The data file KOMMUNE.dta contains all relevant information for this assignment. The file can be downloaded from the course website on Absalon.

Documentation of the data

The data set is a cross-section of 98 municipalities in Denmark in 2010. All nominal variables below are measured in 2010-DKK.

Variable name	Description
Nr	Municipal identifer
Kommune	Name of municipality
Taxrev	Total tax revenues plus other municipal income (in DKK-mio.)
Taxrate	Municipal income tax
Pop	Number of residents

Source: Ministry of Economic Affairs and the Interior, www.noegletal.dk.

Introduction to the assignment

In Denmark, local governments are responsible for primary education, day care services and institutions for elderly. The provision of these local services are financed by tax revenues, transfers from the central government and administrative fees. The municipal income tax is by far the most important source of revenue for local governments and it accounts for up to 70 percent of total municipal revenues.

The municipal income tax rate is set by the local governments themselves and they have full discretionary power over the statutory tax rate as long as it complies with the regulations agreed upon by the Ministry of Finance and Local Government Denmark.

In this assignment, we will study the relationship between total municipal revenues and the municipal income tax using regression analysis.

Problem 1: Descriptive analysis

Present a descriptive analysis of the variables in the file KOMMUNE.dta. Provide one or more tables that present relevant characteristics for each variable. Comment briefly on the results.

Problem 2: Empirical analysis of tax revenues and municipal tax rates

Consider the following regression model and assume for now that it satisfies MLR.1–5:

$$log(taxrev_m) = \delta_0 + \delta_1 \, taxrate_m + \epsilon_m \quad m = 1, ..., 98. \tag{1}$$

1. What is the interpretation of δ_1 in regression model (1)?

2. What is the expected sign of δ_1?

3. Estimate regression model (1) using Ordinary Least Squares (OLS). Report the OLS parameter estimates in a table with their standard errors. Comment on the results. Does the estimate $\widehat{\delta}_1$ match your prediction from Problem 2.2?

Now consider an alternative regression model which includes the local population size as an additional explanatory variable:

$$log(taxrev_m) = \beta_0 + \beta_1\, taxrate_m + \beta_2\, log(pop_m) + u_m \quad m = 1, ..., 98. \tag{2}$$

Assume that model (2) satisfies MLR.1–5.

4. What is the interpretation of β_1 in regression model (2)? In what way is the interpretation of β_1 different than the interpretation of δ_1 in model (1)?

5. Estimate regression model (2) by OLS. Report the OLS parameter estimates in a table with their standard errors. Comment on the results. Does the estimate $\widehat{\beta}_1$ match your prediction from Problem 2.2?

6. The difference between the estimates $\widehat{\delta}_1$ and $\widehat{\beta}_1$ arises from omitting log local population size, $log(pop_m)$, in regression model (1). Explain what omitted variable bias is and describe how the covariance between $taxrate_m$ and $log(pop_m)$ influences the sign and size of the bias in the OLS estimate $\widehat{\delta}_1$.

Problem 3: Multiple linear regression as two simple linear regressions
Assume that the true statistical model is defined as:

$$y = \beta_1 x_1 + \beta_2 x_2 + u \tag{3}$$

where u is an error term.

Note that model (3) does not include an intercept, since the averages of all variables $(\bar{y}, \bar{x}_1, \bar{x}_2)$ are assumed to be zero. This is a simplifying assumption to reduce the algebra involved in the following questions. The basic results are, however, unchanged if an intercept had been included in the statistical model.

1. Write up model (3) using matrix notation. Show that the OLS estimator, $\widehat{\beta}_1$ can be written as:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_{i1} y_i - \frac{\sum_{i=1}^{n} x_{i1} x_{i2}}{\sum_{i=1}^{n} x_{i2}^2} \sum_{i=1}^{n} x_{i2} y_i}{\sum_{i=1}^{n} x_{i1}^2 - \frac{\left(\sum_{i=1}^{n} x_{i1} x_{i2}\right)^2}{\sum_{i=1}^{n} x_{i2}^2}} \tag{4}$$

[Hint: Use $\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = (X'X)^{-1}X'y$]

2. Now consider two simple regressions:

    a. A simple regression of x_1 on x_2 (without an intercept). Define the residuals from this regression as $\widehat{r}_1$.

    b. A simple regression of y on $\widehat{r}_1$ (without an intercept).

    Show that the estimated slope parameter from step b) is identical to $\widehat{\beta}_1$ as defined in equation (4).

    [Hint: Use equation (2.66) in Wooldridge to obtain an expression for the slope parameter in step a). Calculate the residuals based on this expression and use these in step b)]

3. You are now asked to verify your theoretical result from Problem 3.2. First, run a simple regression of $taxrate_m$ on $log(pop_m)$ and an intercept. Save the residuals from the regression. Comment briefly on the OLS results.

    Hint: If you used statsmodels to run your regression, and if you used the variable name 'results' for your OLS results object in Python, you can use the code below to save the residuals from the regression as a new variable called 'res1' in your DataFrame:

    ```
    df['res1'] = results.resid
    ```

    Next, run a simple regression of $log(taxrev_m)$ on $res1_m$. Compare the parameter estimate to what you obtained in Problem 2.5.

    One interpretation of the two-step procedure is that the first step divides the variation in $taxrate_m$ into two components: $taxrate_m = \widehat{taxrate}_m + \widehat{res1}_m$. Discuss why it is possible to obtain an estimate identical to $\widehat{\beta}_1$ in the second step when $log(pop_m)$ is an omitted variable.

    [Hint: Does the second step suffer from omitted variable bias? Why? Why not?]

Problem 4: Conclusion
What are your overall conclusions regarding the relationship between total municipal revenues, the income tax rate and local population size?