

# Microeconometrics: Self-Studies

## 1 Organisation

Students form teams up to 4 members to work on the project together. All team members agree in the beginning to the group composition. Grading of the team output is independent of the number of members. Each team elects a ‘captain’ who is in charge of communication. Each captain informs us not later (preferably earlier) than the 2<sup>nd</sup> Friday after the break, i.e. no later than **30.04.2021** about the team member composition (email to: [gabriel.okasa@unisg.ch](mailto:gabriel.okasa@unisg.ch)). Immediately after receiving this information, we will provide the data necessary to start working on the project & provide digital joint working space on Switchdrive or Canvas, if requested. Everybody has to be part of a team after 30.04.2021 or is considered to be a single-member team. In the beginning of each lecture (i.e. Q&A session), we spent up to 15 min with your questions about the self-studies project.

## 2 Task

The task is to conduct an empirical study from the beginning to the end. The research question is as follows:

***‘Money does not buy success’. Is this true? In a soccer game, how many additional points per game can a home team expect if it is more expensive than the team of the visitor?***

Each group will receive a (different) random sample of the data set with games from the German soccer league (Bundesliga 1) & documentation of the data set at hand. There are few things to consider:

- **Identification:**

- It is up to you to choose (& defend) an appropriate identification strategy

- **Estimation:**

- It is up to you to choose an estimator which you consider appropriate and code such estimator in Python.
- Restrictions:
  - Use at least two different estimators & compare results (& explain what you learn from this comparison).
  - At least one of the two estimators should be among those discussed in the lectures.

- **Coding:**

- Full points only for implementation in Python.
- Support is provided only for Python (which is the highly recommended option). Gauss is accepted as well.
- Implementation in R & Stata is tolerated but it is almost impossible to receive full points (only if substantial advanced coding is contained in the solutions, ideally in C++ embedded in R using Rcpp or similar).
- You get Python code throughout the lecture/tutorials (as well as in Data Analytics 2). Code is also available from other sources. You are allowed to reuse code with proper detailed reference, e.g., «Taken from Source XXX; downloaded on <date>». You are also allowed to adapt the code with the reference «Adapted from Source XXX; downloaded on <date>». However, writing everything in your own code, receives substantially more points. The more independent the coding, the better.
- Projects of different groups that are (almost) identical will receive **zero** points.

- **Documentation:**

- The documentation should not exceed 15 DIN A4 pages in a standard format.
- Structure of documentation:
  - Describe your research question & your data.
  - Describe your parameter of interest, your identification strategy & discuss its plausibility.
  - Explain the estimators under consideration & why you choose it.
  - Compactly summarize your results in few (!) tables or graphs & discuss your findings (What have we learned? Is this in line with the expectations? Robustness of results? ...)

- **Warning:**

- For each group, there is a lot to do & not much time!
  - Lots of coding (& testing, debugging etc.).
  - Lots of data analysis.
  - Lots of freedom which estimators to use & how to implement them.
- This requires careful planning for each group!
- Split tasks equally among all team members.
- Add a detailed explanation who has done what (signed by all members).
- The team captain (or somebody else) should make sure that the ‘production runs smoothly’ & internal deadlines are met.

### 3 Submission

The submission deadline is the day before the presentation, until midnight, i.e. until Tuesday **25.05.2021 23:59**. A complete submission contains the following parts:

- A max. 15 page pdf file with documentation.
- A file with the code:
  - documented, i.e. easily understandably by another person.
  - executable (Python, Spyder), i.e. replication must be possible.
- A presentation (pdf or PowerPoint).
- An explanation of which group member did what.

### 4 Presentation

Results will be presented on the last Wednesday in the term (after the last PC Lab session), i.e. on **26.05.2021**. Each presentation will be up to 30 min (maybe shorter depending on the number of groups that will present – information will follow once the group compositions are finalized). Every team member must present to get credits. Present all parts of your empirical research, i.e. from data cleaning to the implications of the findings. Adapt the technical level of the presentation to the target audience of students who have passed Data Analytics 1, Data Analytics 2, and this course.

### 5 Evaluation

The total of **25 points** is decomposed as follows:

- **10 points** for the documentation:
  - show that you understand the setting, are careful with the data, make reasonable choices w.r.t. identification & estimation, and that you can interpret the results correctly.

- **9 points** for the implementation consisting of the product of the following two parts:
  - ***Ambition*** (1-3 points):
    - 1 point: mostly reused available code and packages (Python, Gauss, R, MATLAB, Stata)
    - 2 points: mix of reused, adapted and new code (Python, Gauss, R, MATLAB, Stata)
    - 3 points: mostly new code (Python, Gauss, MATLAB)
  - ***Programming*** (0-3 points):
    - 0 points: if replication not possible.
    - 1 point: if the code runs correctly and replication is possible.
    - 2 points: if the code is well-commented and well-structured.
    - 3 points: if the coding is efficient.
  - *For example*, a code with 2 points for the ambition & 3 points for programming gives 2 x 3 points, i.e. 6 points in total.
- **6 points** for the presentation.