Tips for writing a Bachelor/Master thesis

Sebastian T. Braun

Overview

Formulating a research question

Structure
- Introduction
- Literature Review
- Empirical Strategy
- Data
- Results
- Conclusion

Writing tips

Additional tips

Resources

Choosing a topic

In choosing a topic, you should consider the following issues:

1. **Motivation**: Choose a topic that you find interesting.
2. **Literature**: Choosing a topic with a large literature is often a good idea, as you can build on previous work.
3. **Feasibility**: Appreciate your time and resource constraints.
4. **Originality**: What will be your point?

Formulating a research question

- Avoid spending too much time on broad, vague problems.
- Instead, focus on **specific questions** early on.

What is a good research question?

1. **Important**: The ’So what?’ question.
2. **Specific**: Narrow down your focus.
3. **Answerable**: The question should be answerable within your constraints.
4. **Innovative**: What will be your point?

What is a good research question? Example

- **Bad**: What happens to native workers when there is a lot of immigration?
  - Important, but not specific. Might not be answerable (within your constraints).
- **Good**: What was the effect of Cuban immigration to Miami in the aftermath of the Mariel Boatlift on wages of non-Cuban low-skilled workers? (Card, 1990)[1]

[1] Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *ILR Review* 43(2): 245-257.

Framing a hypothesis

A **hypothesis** is a potential answer to a research question.

Why is it a good idea to formulate hypotheses?

- Having a potential answer in mind narrows the scope of your project to a specific aspect of a research problem.
- Hypotheses tell you what data to collect, what ingredients your model should have, or how to design an experiment.
- Hypotheses based on economic theory make your empirical results more convincing.

What is a good hypothesis?

A good hypothesis should:

1. **Answer** your question.
2. Be **specific**.
3. Be **falsifiable**. There should be a reasonable 'alternative' hypothesis.

An example from my own research

Observation: Unemployment rate of forced migrants in post-war Germany differed greatly across West German regions.

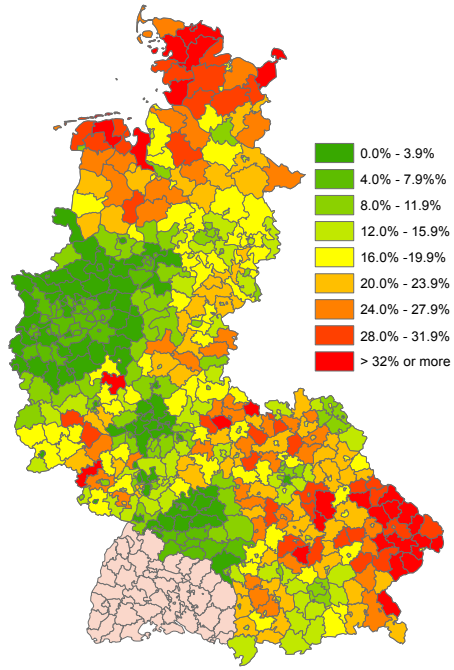


Abbildung 1: Unemployment of forced migrants in W Germany, 1950

An example from my own research (ctd.)

Observation: Unemployment rate of forced migrants in post-war Germany differed greatly across West German regions.

- **Q**: Why did unemployment rates of forced migrants in post-war Germany differed greatly across regions?
- **H**: Higher immigrant inflows increased the unemployment rate of forced migrants.
- **Q'**: Did higher inflows of forced migrants increase the unemployment rate of forced migrants in post-war Germany?

An example from my own research (ctd.)

In this example, framing a hypothesis is useful:

- To narrow down the research question.
- To clarify minimum data requirements (data on regional migrant shares and migrant unemployment).
- To base the empirical exercise on economic theory (if the hypothesis can be rationalized by an economic model).

An example from my own research (ctd.)

In this example, a bad hypothesis may read:

- Regional unemployment rates of migrants will vary with regional immigrant inflows (**not specific**).
- Higher immigrant inflows have no effect on the unemployment rate of native workers (**does not answer your question**).
- Higher immigrant inflows increase, decrease or leave constant migrant unemployment (**not falsifiable**).
- Higher immigrant inflows increase the number of immigrants (**no reasonable alternative**).

Structure–General remarks

- Don’t write your thesis like a crime story–with the (surprising) punch line coming last.
- Instead, figure out your **one central contribution** and put this contribution right at the **beginning** of your thesis.
- Your readers are busy and impatient–help them to get your contribution quickly!

Example: Chetty et al. (2014)[2]

'The United States is often hailed as the 'land of opportunity,' a society in which a child's chances of success depend little on his family background. Is this reputation warranted? We show that this question does not have a clear answer because there is substantial variation in intergenerational mobility across areas within the U.S. The U.S. is better described as a collection of societies, some of which are 'lands of opportunity' with high rates of mobility across generations, and others in which few children escape poverty.' (Page 1554)

[2]Chetty, R., N. Hendren, P. Kline and E. Saez (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *QJE* 129(4): 1553–1623.

Structure–Overview

The typical structure of a paper in quantitative economic history is:

1. Introduction
2. Literature Review
3. Historical Background
4. Data
5. Empirical Strategy
6. Results
7. Conclusion

Structure–Overview (ctd.)

Of course, there are **many variants** of this structure:

1. The Literature Review is often part of the Introduction.
2. You will sometimes see a separate Discussion section.
3. The Data section often precedes the Empirical Strategy.
4. (...)

The structure of your thesis might well look different!

Introduction

The Introduction should be **brief and concise**–and **motivate** the reader to read the rest! What should be in it?

- Which question/hypothesis do you address/test?
- How do you answer the question?
- Why should we care about this question?
- What is the contribution of the paper?
- What are the main results and what are their implications?

Example: Butcher and Card (1991)[3]

Introductions often progress through four 'moves' which can appear in different order (Dudenhefer 2014):

1. Establish a **research territory**: Introduce your subject and indicate its importance.

   'More immigrants entered the United States during the past decade than in any comparable period since the 1920s. Among the issues raised by this influx, none is as controversial as its effect on the labor market opportunities of native-born workers.' (Page 292)

[3] Butcher, K.F. and D. Card (1991). Immigration and Wages: Evidence from the 1980's. *AER* 81(2): 292-296.

Example: Butcher and Card (1991) (ctd.)

2. Point towards **literature**: Briefly point towards the literature.
3. **Identify niche**: Point out potential gaps and uncertainties in the literature.

   'Evidence on the labor market consequences of immigration is limited (see Greenwood and McDowell 1986 and Borjas 1990).' (Page 292)

Example: Butcher and Card (1991) (ctd.)

4. **Occupy niche**: Describe how you address these deficiencies and what you find.

   'This paper presents new evidence on the effects of immigration, based on changes in the distributions of wages in 24 major cities during the 1980s. Our empirical analysis reveals large differences across cities in the relative growth rates of wages for low- and high-paid workers. Nevertheless, these differences bear little or no relation to the size of immigrant inflows. Our results therefore confirm the findings of earlier studies [...] that suggest that the labor market consequences of higher immigration are relatively small.' (Page 292)

How to begin the Introduction?

There are many ways to begin your thesis. Here are some examples from the *AER*:

- State **what you do**: 'This paper evaluates the welfare effects of corporate income tax cuts on business owners, workers, and landowners.' (Serrato and Zidar, 2016: 2582)
- State relevant **background**: 'The age of mass migration from Europe to the New World was one of the largest migration episodes in human history.' (Abramitzky et al., 2012: 1832)
- State an **economic question** or fact: 'What should the marginal tax rate on top income earners be?' (Ales and Sleet, 2016: 3331)

Why undertaking a literature review?

- To increase your **breadth of knowledge** and demonstrate your familiarity with academic work on the topic.
- To **avoid re-inventing** the wheel.
- To **identify gaps** or shortcomings in the literature and **motivate** your work.
- To be able to build on existing work and put your results into **perspective**.

What should a good literature review do?

A literature review should NOT just summarize, in no particular order, a series of studies on a topic. Instead, it should:

- Analyze **critically**, and **organize**, a body of research.
  - A good review has structure.
- Provide **context and motivation** for your own research.
  - A good review describes how your work builds upon and extends the existing literature.

Suggestions for writing a review

- Begin with comments about the **body of research as a whole**.
  - Are there many studies or few? Are they mainly empirical or theoretical?
  - Is there a general consensus on certain issues? Is there disagreement on others?
  - (...)
- **Organize your review** according to themes (methodology, conclusion, data,...), not by individual papers.

Suggestions for writing a review (ctd.)

- **Begin paragraphs** with a sentence that puts in **explicit context** what follows.
  - Don’t leave it to the reader to infer the point you are making.
- **Critically discuss** merits and shortcomings of the literature.
  - Be explicit!

Suggestions for writing a review (ctd.)

- Explain how your work **builds upon and adds** to the existing literature.
  - Do you have new data?
  - Do you add a twist to a model?
- Your contribution is usually a direct response to a shortcoming identified in the review.

Example: Kerr and Kominers (2015)[4]

'Research on the spatial horizons over which different agglomerative forces act often takes one of two approaches. A first approach considers regional evidence. (...) In parallel, a second strand of work considers local evidence on agglomerative interactions. (...) There is a substantial gap between these two approaches. (...) This project examines these issues theoretically and empirically.' (Page 877)

[4] Kerr, W.R. and S.D. Kominers (2015). Agglomerative Forces and Cluster Shapes. *RESTAT* 97(4): 877–899

Empirical Strategy

The Empirical Strategy section **discusses the empirical methods** you use to test your hypothesis. What should be in it?

- Describe the empirical methods you will use (OLS, 2SLS,...).
  - Assume your readers know the basics of the method.
  - Do not assume they know how you apply the specific method in your context.

Empirical Strategy (ctd.)

- Write out the basic econometric specification and explain each variable and the parameters of interest.
  - Was the equation derived from theory? Has it been used before?
  - Why do you include certain variables (and not others)?

Example: Boustan et al. (2010)[5]

IV. Estimating the Relationship between Internal Migration and Economic Outcomes

A. Ordinary Least Squares Specification

We are interested in the effect of internal migration to a metropolitan area on the employment rate, annual earnings, and out-migration rate of existing residents. Let $Y_{ijr\,40}$ represent an economic outcome for a nonmigrant i who lives in metropolitan area j in region r in 1940. We posit that $Y_{ijr\,40}$ will be a function of the migrant-induced change in labor supply:

$$Y_{ijr40} = \alpha + \beta m_{jr40-35} + \gamma o_{jr40-35} + \Phi' Y_{jr35} + \Psi' X_{ijr40} + \Pi_r + \varepsilon_{ijr40}, \quad (1)$$

where $m_{jr\,40-35}$ represents the in-migration rate to area j and $o_{jr\,40-35}$ represents the out-migration rate from area j. In the earnings and employment equations, we expect β to be negative and γ to be positive. The most

[5]Boustan, L.P., P.V. Fishback and S. Kantor (2010). The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression. *JOLE* 28(4): 719-746.

Empirical Strategy (ctd.)

- If you wish to interpret a parameter as causal (here: β), critically discuss the assumptions you need to make.
  - It is far better to acknowledge potential problems than to make unsupported statements.
- Describe additional specifications or tests you will perform.
  - Alternative samples.
  - Additional control variables. Alternative variable definitions.
  - Alternative regressions methods. (...).

Data

The Data section is usually short. What should be in it?

- **Identify data source(s)**: Where do the data come from?
- **Describe data source(s)**: Periods covered; number and unit of observations; panel, cross section or time series; (...).
- Describe **limitations and strengths**: Missing observations; sample attrition; representativeness; (...).
- Describe construction of **sample and variables**: Deleted outliers? Combined two variables to get a third? (...).

Data (ctd.)

The second part of the Data section presents **descriptive statistics**.

- Report at least the mean and standard deviation of variables, and the number of observations.
- For categorical data, report the percentage of observations in each group.
- You may want to distinguish between outcome, independent and control variables or between different sample subgroups.
- Discuss the most relevant variables–you don't have to discuss every single mean!

Useful questions

- How does your data compare with other data sources used in the literature? Might help you to identify your contribution.
- What would the ideal data look like? Might help you to identify weaknesses.
- Are there any data features that obviously affect your results? Might inform you about robustness checks or future work.

Results

There are essentially **two decisions** to make when presenting your results (Neugeboren and Jacobsen, 2014)[6]:

1. How many of your results should be presented (and which)?
2. How should these results be described?

[6] Neugeboren, R. and M. Jacobson (2014). *Writing Economics*. Harvard Writing Guides.

How many results should I report?

- **Less is usually more**.
- Pages of parameter estimates etc. muddy the message of your thesis.
- Present those estimates that directly speak to your topic.
  - If it's not worth writing about in the text, it's probably not worth putting in a table or figure.

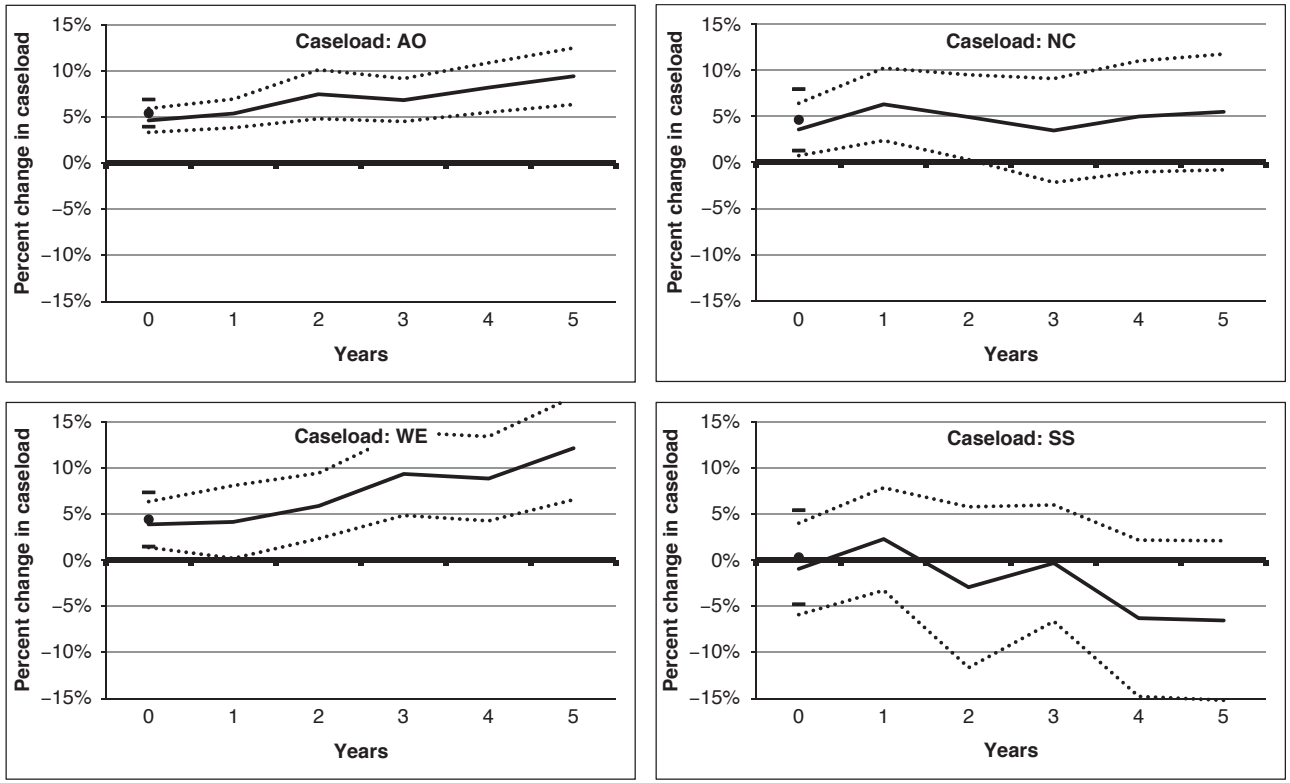
Visualizing data

- Good figures really make a paper come alive.
- Three basic principles (Schwabish, 2014)[7]:
  1. **Show and emphasize the data**. The data are the most important part of the graph.
  2. **Reduce clutter**. Unnecessary or distracting visual elements tend to reduce effectiveness.
  3. **Integrate the text and the graph**. Integrate legends. The figure should contain enough information to stand alone.

[7] Schwabish, J.A. (2014). An Economist's Guide to Visualizing Data. *JEP* 28 (1): 209-34.

An original line chart (Schwabish, 2014)

Figure 1A
An Original Line Chart



Source: Klerman and Danielson (2011).

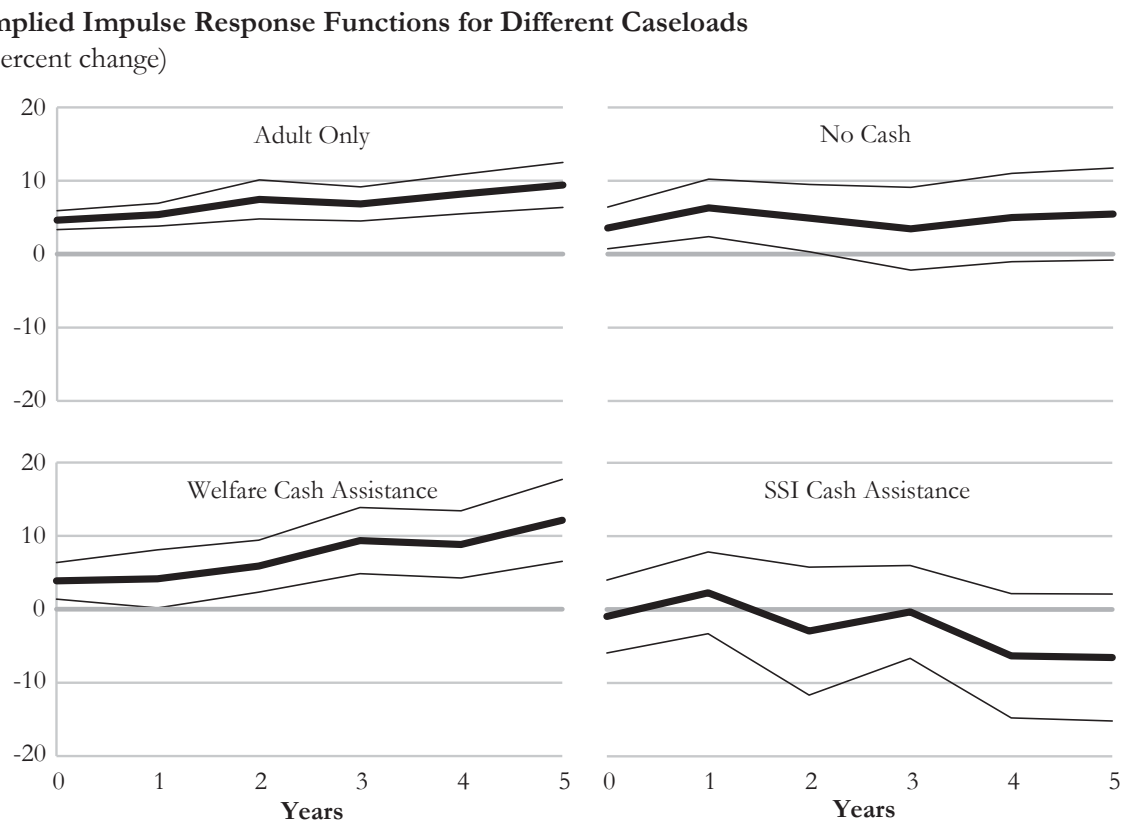
An original line chart (Schwabish, 2014) (ctd.)

The graphs violates the three principles outlined above:

- The graph does not emphasize the data:
  - Your eye is drawn to the horizontal gridline.
- Some elements add unneeded clutter.
  - The y-axis labels and percentage signs are redundant.
- The figure does not stand alone.
  - What do AO, NC, WE, and SS mean in the figure?

A revised line chart (Schwabish, 2014)

Figure 1B
A Revised Line Chart



A revised line chart (Schwabish, 2014) (ctd.)

- The darkest line now shows the data.
- Two sets of labels have been eliminated
- The title now identifies the unit; the percent signs in the figure have been deleted.
- Repositioning the word ’Caseload’ into the title leaves room to spell out the abbreviations.

How should I describe my results?

- Describe your tables, graphs and figures **clearly and precisely**.
- Introduce a table (figure) and identify the main points made by the data in the table (figure).
  - You cannot simply refer to a table and leave it at that.
- **Guide readers** and focus their attention on the important parts of the table, in the right order.
- Establish the main point of the table in the **topic sentence** of a paragraph.

Example: Neugeboren and Jacobsen (2014)

'Table I shows that including a measure of ability in the wage equation dramatically lowers the predicted effect of education on earnings. Column 1 does not include an ability measure and indicates that a year of education raises wages by 9.1 percent. Column 2 adds the ability measure and the education effect drops to 3.1 percent. Columns 3 and 4 show that this general pattern is repeated even when state level dummy variables are included. The estimates in Table I are therefore consistent with the hypothesis that the OLS estimates suffer from an upward ability bias.' (Page 27)

Example: Neugeboren & Jacobsen (2014) (ctd.)

- **Focus** on the parameter(s) of interest.
- The reader should be able to understand the table **without having to look back at the text**
  - Define the dependent variable.
  - Do not report your estimates in terms of α and β, give variables informative names.
  - Add table notes.
- Don't report 1.23456789.
- After presenting your main results you may want to discuss additional robustness check(s).

How should I describe my results? (ctd.)

- Discuss your results as **honestly** and carefully as possible.
  - Tell readers if your results do not sit well with your hypothesis.
- Explain the **economic significance** of central numbers, not just their statistical significance.
  - Explain what the numbers mean. Provide units.

Conclusion

The Conclusion is **brief**. It usually looks both **back at a paper and ahead of it**.

- Briefly restate your principal contribution and finding.
- Briefly discuss the wider implications of your thesis.
  - Has your finding policy implications? Is it relevant beyond your specific context?
  - Avoid normative judgments. Don’t speculate. Rely instead on economic analysis.

Conclusion (ctd.)

- Briefly identify limitations of your project. You might also suggest avenues for future research.
  - Recall: It is way better to acknowledge limitations than to make unsupported claims.

Example: Kline and Moretti (2014)[8]

The Conclusion basically proceeds in three steps:

- Step 1: Restate main contribution.

  'This article makes two primary contributions. Our substantive contribution is to estimate the local and aggregate effects of one of the largest place-based policies in U.S. history. To our knowledge, we are the first to empirically quantify the long-run social costs and benefits of such a policy. [...]' (Page 327)

[8] Kline, P. and E. Moretti (2014). Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *QJE* 129(1): 275–331.

Example: Kline and Moretti (2014) (ctd.)

- Step 2: Discuss policy implications.

  'Our empirical findings are policy relevant. The evaluation design of Section III provides strong evidence that the TVA sped the industrialization of the Tennessee Valley and provided lasting benefits to the region in the form of high-paying manufacturing jobs. [...] (Page 328)

Example: Kline and Moretti (2014) (ctd.)

- Step 3: State limitations. Discuss future research.

  'We caution, however, that our findings do not necessarily apply to all contexts, as the strength and shape of agglomeration economies may well vary across industries, periods, and levels of aggregation. Our results are specific to the manufacturing sector and a period of U.S. history when manufacturing employment was expanding and earnings were relatively high. An important task for future work is to assess whether similar qualitative results hold for modern development efforts, such as those centered on building high-tech clusters.' (Page 328)

Writing tips (Dudenhefer, 2014)

- Use the active voice.
  - X In this paper, the effect of immigration on native wages is studied.
  - ✓ This paper studies the effect of immigration on native wages.
- Express key actions as verbs.
  - X There is *opposition* among many voters to nuclear power plants.
  - ✓ Many voters *oppose* nuclear power plants.

Writing tips (ctd.)

- Keep your grammatical subjects short.
  - X An explanation of why the model cannot accommodate this particular case of omitted variables bias is given in the appendix.
  - ✓ The appendix explains why the model cannot accommodate this particular case of omitted variables bias.
- Avoid unnecessary words.
  - Change ’is equal to’ to ’equals’.
- Keep sentences short. Be specific.

Writing tips (ctd.)

- Avoid adjectives and verbs that are overly dramatic.
  - 'The results shatter our expectations' is too much.
- Use 'I' when you mean 'I' and use 'we' when you mean 'we'.
- It may seem boring to keep using phrases like 'The results show...', 'The estimated coefficient on...'. Use them anyway.
  - Use them anyway or use anything equally clear.
- Revise and polish. Revise and polish. Revise...

Additional tips

- Introduction and Conclusion are usually written at the end of your project.
- The more economics papers you read, the more familiar you will become with their structure, conventions and style.
- You may not reach conclusive results. It is better to acknowledge shortcomings than to make unsupported claims.
- Rejecting your hypothesis is totally fine (as long as the hypothesis is reasonable to begin with).

Additional tips (ctd.)

- Don’t underestimate the time to organise data and write-up research.
- Time management is key.
- Back up often and in multiple places.
- Plan the meeting with your supervisor.
- Get a smart friend to read your proposal/thesis.
- Never ever plagiarise.

Resources

...which I have used to prepare these slides.

- Dudenhefer, P. (2014). *A Guide to Writing in Economics*. Duke University. (available here)
- Neugeboren, R. and M. Jacobson (2014). *Writing Economics*. Harvard Writing Guides. (available here)
- Nikolov, P. (2013). Writing Tips For Economics Research Papers. *Mimeo*, Harvard University. (available here)
- Schwabish, J.A. (2014). An Economist's Guide to Visualizing Data. *JEP* 28 (1): 209-34.
- Zinnser, W. (2016). *On Writing Well: The Classic Guide to Writing Nonfiction*. Harper Perennial.