Guidelines for the resesarch proposal

Nurfatima Jandarova

Content

The proposal should include

- a clear and specific research question
- motivation for research question
- contribution to existing works

Specific research question

Generic form of research question is "How does X affect Y?". This is how we typically start thinking about any question. For example, "How do student benefits affect their outcomes?".

However, as a research question, this is too vague. Trying to answer a vague research question can quickly become overwhelming. There may be different types of benefits that students receive, each with different amounts and different conditions. There are, definitely, myriads of student outcomes: do they work while studying, what kind of grades they receive, how many classes they attend, how they fare in exams, do they do summer internships, do they change fields or universities, what kind of jobs they find after graduating, how fast do they find jobs after graduating, etc.

You can transform this vague research question into a more specific one by narrowing down the X and Y as much as possible. A more specific version of research question in the example could be "How does reduction in housing allowance affect graduation timing of students?" or even better "How does the transition from housing allowance to housing supplement affect graduation timing of students?"

When you formulate a sufficiently specific research question, it makes the following steps much easier to think about!

Important research question

Once you have your research question, you should also demonstrate its importance.

- **Why answering your RQ is important?**
- **How could the society benefit from the answer?**

Economics is all about trade offs. It can help to identify the key tradeoff that you want to focus on. In the example with housing allowances, the tradeoff could be that government is saving money right now by cutting the allowance, but can lose money in the future if these students end up working in low-paying jobs (maybe they drop out of universities, or they don't have time to study properly because they have to work more to pay rent). You can think how to value each side of the tradeoff. In our example, they could be measured in EUR terms as reduction in government expenses today and reduction in government tax revenues in the future. The value that you attach to the tradeoff depends on the question you want to study. Not everything needs to be measured in money terms, but it might be easiest way in many cases.

A tip: you can look how other papers on similar topics motivated their projects.

Novel research question

Finally, your specific and important research question should offer us something novel.

- It can be a completely new evidence never studied before (which is extremely rare).
- It can be an old question studied before, but looked at from a new angle.

In either case, you need to demonstrate how this question (or related topics) have been studied before. Finding papers should not be difficult in our day and age. But finding reputable papers can be sometimes tricky.

If your topic is covered in this course, reading list is a good place to start. Even if none of the papers are exactly on the topic you've chosen, you can use a closely related paper and find references to more relevant studies in their literature section.

The papers in the reading list also give you an idea about trustworthy journals. You can use this information to filter Google Scholar search results.

Make sure that you explicitly state how your project can offer new evidence! Do not simply list past papers and what they've done. You need to explain how your proposal is different from what they've done and why this difference is important to account for.

In the example with housing allowance, there are plenty of papers that have studied before financial assistance to students.

- You can claim that this question has never been studied in Finland and that it was the only country to have generous allowance before the policy change (may not be true!). So, it would give you a unique opportunity how reduction from generous to minimal assistance affects students.
- Alternatively, it might be that previous papers could not separate between other types of benefits and housing allowance, but you found new data source where this distinction is possible. And again, need to explain why separating housing allowance from other benefits is important in understanding the economic tradeoff of students.
- If it turns out that all the other papers were studying benefit expansion, you could argue that response of students to benefit reduction may be not be symmetric.

- It can also be that all the other papers used general surveys that primarily cover middle-class families and do not represent well very poor students. You could then argue that the policy change would be a lot more painful for poor students whose families cannot offer substitute housing. So, their budget constrain would look different from middle-class students and they'd make different decisions. In this case, make an attempt to verify that data with good representation of poor students exists and can be accessed/bought. **You do not need to access/buy any data!** Only verify existence, sample and scope.

Overall, you need to 1) understand current literature on the topics and 2) explain your contribution relative to existing studies.

Main body

The rest of the guideline depends whether you want to make a **theoretical** proposal or an **empirical** one. Both proposal types may include elements of the other. For example, a theoretical proposal may include discussion of how to test the theory with the data. Or an empirical proposal may include discussion of a theoretical framework that could help make sense of the estimates.

Theoretical proposal

- Clear description of model components and parameters
  You should clearly describe
  - who is an agent making a decision
  - what kind of information she observes
  - what decisions does she make
  - what are the constraints she is facing
- Simple predictions, comparative statics or simulations
- Bonus: outline of the plan to estimate model parameters with the data

Empirical proposal

- Description of a potential dataset
  The question you want to answer here is why this dataset is suitable to study the question at hand.
  - *Sample members*: does the dataset include the people whose behaviour you would like to study?
    For example, if you want to study minimum wages in the fast food industry, you need to have sufficient amount of workers in that industry and maybe similar people that work elsewhere or don't work at all. Can the chosen dataset contain information about these people? You may not find exact numbers without actual access to the data, which is ok. But try to give as best an answer to this question as you can.

  - Key variables: does the dataset include the variables that you would like to study?
    For example, does it include information whether people were working or not, which industry they worked in, maybe the occupation they had? This information should be available in the documentation of any dataset, so you should be very precise about it. Discuss if the variable measures the quantity of interest directly or if you need to apply some transformations to the variables before you get a proxy for the quantity of interest? For example, many datasets do not include hourly wage information, but have monthly earnings and typical hours worked in a week. You could compute proxy of hourly wage by dividing the monthly earnings by hours worked in month. How could this transformation affect your analysis?

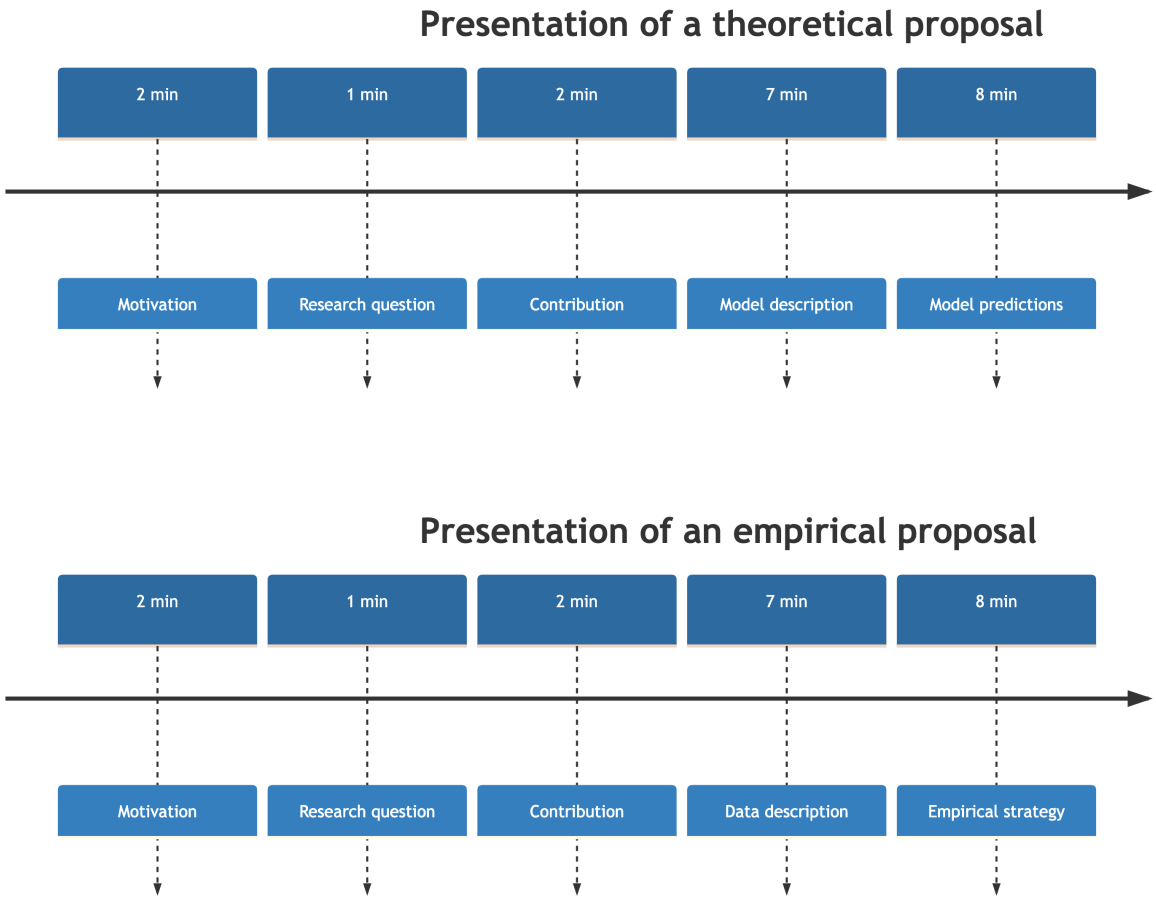
- Discussion of empirical strategy
  - Clear statement of empirical strategy.
    For example, instrumental variables, or regression discontinuity, etc.
  - Why do you think this method is well suited to answer your research question?
  - Clear definition of necessary components.
    For example, an instrument variable for IV, or running variable for RDD.
  - Discussion of assumptions required for the chosen method.
    For example, exogeneity and relevance for IV. You should discuss why you think chosen instrument may be relevant instrument and what could endanger exogeneity of the instrument.
- Bonus: discuss how the estimation results may be interpreted in the context of a relevant economic model.

Time structure of the presentation

Aim to spend first 5 minutes on motivation, statement of research question and potential contributions.

Rest 15 minutes should cover the details of the proposal.

Here are example time breakdowns of a 20 minute proposal presentation.



Examples

As I have mentioned at the beginning of the class, best examples are already published papers. Feel free to find your own papers to use as a template.

Below I provide two papers and point relevant parts demonstrating content points above.

Example for a theoretical proposal

Example based on Autor, D. H., Levy, F. and Murnane, R. J. "The Skill Content of Recent Technological Change: An Empirical Exploration*," *The Quarterly Journal of Economics* 118, no. 4 (November 1, 2003): 1279–1333.

Research question

> "what it is that computers do— or what it is that people do with computers—that causes educated workers to be relatively more in demand." (p. 1280)

Motivation

The authors give rather short motivation and contribution because the work was quite novel. The main motivation is by stressing the "gap" in the literature: previous literature documented positive correlation between computerisation and increased employment of educated workers without providing the causal framework.

Read first paragraph of the Introduction on pages 1280-81.

Contribution

The contribution is essentially given by the motivation. Once you identify the gap in the literature, filling that gap is usually the contribution. The authors provide some discussion of possible contributions in footnote 2 at the bottom of page 1280.

Model description

All the elements, underlying assumptions and rationale for those are discussed in detail in subsection I.A. on pages 1286-1289.

Model predictions

Some predictions are given in the last three paragraphs of subsection I.A. on page 1289. Industry predictions are described in subsection I.B. on pages 1289-91.

It is part of your job to identify what predictions from the model are of interest. You do not need to generate predictions in terms of every single input into the model. Choose one or two that are central to answering the research question.

Bonus: estimation in the data

Entire section II on pages 1291-95.

Example for an empirical proposal

Example based on Kleven, H., Landais, C. and Søgaard, J. E. "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics* 11, no. 4 (October 2019): 181–209.

Research question

> "the effects of children on the careers of women relative to men" (p. 181)

Motivation

First paragraph on page 181. The authors relate to a larger debate about gender pay gap and how despite convergence over the years, it is still not eliminated. They propose their hypothesis about the role of childrearing as a possible explanation.

Contribution

The contributions are discussed at length on pages 183-184. Pay attention to sentences that highlight the differences of the current paper relative to those that have already been published such as last sentence of the first paragraph on page 184.

Data description

Subsection I.B. on pages 186-87. Pay attention to sample selection criteria described in the second paragraph of the subsection. Many of the key variables are simply listed in the first paragraph because they are self-explanatory and their definitions are standard. However, pay attention to the discussion of construction of hours worked and wage rate variables. Especially, how the data limitations in their definition may contribute to biased results.

Empirical strategy

Subsection II.A on pages 187-88 provides description of the chosen methodology. Notice how careful the authors are in defining different elements of the regression equation.

Identification paragraphs in subsection II.B. on pages 195-96 discuss at length necessary assumptions for their empirical strategy to work as well as outline of robustness checks to strengthen their case.

Bonus: mechanisms

Section IV on pages 201-04.

References

Autor, David H., Frank Levy, and Richard J. Murnane. "The Skill Content of Recent Technological Change: An Empirical Exploration*." *The Quarterly Journal of Economics* 118, no. 4 (November 1, 2003): 1279–1333.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11, no. 4 (October 2019): 181–209.