

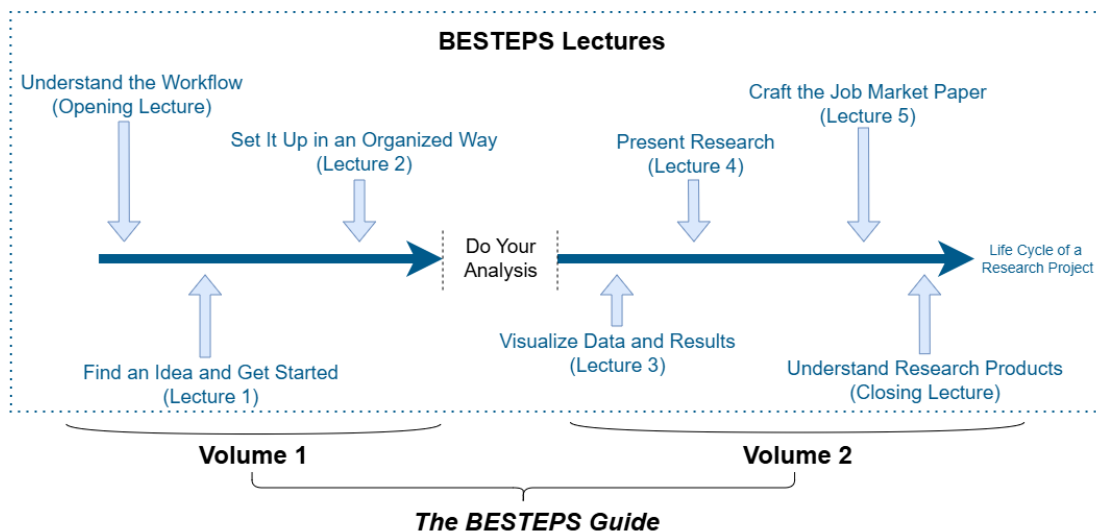
The BESTEPS Guide:

7 Lessons towards Effective Economics PhD Research

Edited by: Anna Chesa-Llorens* and Mengwei Lin 林孟蔚†

The BESTEPS Guide is intended to offer early-career economists, especially PhD students, insights from leading researchers on how to take the “best steps” at each key stage of a research project. It can be used for self-learning or for courses on research skills.

The *Guide* is based on a series of lectures held at the University of Barcelona (UB) School of Economics under the BESTEPS project, delivered by **Antonio Cabrales, Dirk Foremny, Albrecht Glitz, Libertad González, Joan Llull, Guillermo Marin, Diego Puga, and Jenifer Ruiz-Valenzuela**. Each lecture is self-contained and can be read independently. Together, they offer practical and comprehensive guidance on conducting research effectively during the PhD.



*University of Barcelona & Barcelona Institute of Economics (IEB). annachesa@ub.edu.

†University of Barcelona & Barcelona Institute of Economics (IEB). mengwei.lin@ub.edu.

Acknowledgements

BESTEPS (*Barcelona Essential Skills Training for Economics PhD Students*) is a flagship PhD training project organized at the UB School of Economics and funded by the UB Doctoral School. It is a core component of the training framework of the PhD in Economics program at the University of Barcelona.

The lecturers generously agreed to publish these notes as a public good, but did not review the draft. We are grateful to the following young economists across fields and from various institutions for carefully reading the draft and providing thoughtful feedback (in alphabetical order): Bastian Bruestle (Zurich), Haochi Chen (Dublin), Zhenghao Chen (UCL), Alessandro Ciardo (Barcelona), Britta Hecker (IAB), Guillermo Martínez (CEMFI), Taoyi Sun (Purdue), Robert Włodarski (Cambridge), Jessie Ziyue Zhu (Maastricht). All remaining errors are our own responsibility.

Contents

Preface	vi
Volume 1	viii
1 Opening Lecture: Navigating the Research Process in Economics	1
1.1 Introduction: Research as a Power Law	1
1.2 Step 1: Idea Generation and Evolution	2
1.3 Step 2: Choosing the Right Methodology	5
1.4 Step 3: Writing and Communicating	6
1.5 Challenges along the Way	8
1.6 Bonus: Preparing for the Job Market - It is too early, and it is not too early	9
2 Lecture 1: Starting a Feasible and Promising Project	11
2.1 Step 1: Select a Field (or two)	12
2.2 Step 2: Select a Topic	12
2.3 Step 3: Find and Evaluate a Research Question	13
2.4 Step 4: Develop a Plan - Approach and Literature	15
2.5 Step 5: Seek Feedback and Refine the Idea	17
2.6 Step 6: GO! Move Forward with Your Project	18
2.7 Bonus: Do's and Don'ts for Junior Researchers	18
3 Lecture 2: Doing Organized and Replicable Research	20
3.1 Introduction: Reproducibility and Replication Packages	20
3.2 Part 1: Managing Data	23
3.3 Part 2: Coding Organized Projects	27
3.4 Bonus: Main Takeaways and Ten Simple Rules	35
Volume 2	37
4 Lecture 3: Visualizing Economic Research	39
4.1 Part 1. Introduction to Visualization	39

4.1.1	Visualization: Power and Pitfalls	40
4.1.2	Understand Visual Encoding	43
4.1.3	Best Practices for Academic Visual Design	45
4.2	Part 2. Visualization for Economists	46
4.2.1	Data Visualization for Economists	46
4.2.2	Visualize Data Before Estimating	48
4.2.3	Visualize Econometric Results	54
5	Lecture 4: Academic Talks in Economics	60
5.1	Start Presenting Now	61
5.2	Before the Talk: Design Slides and Rehearse	62
5.2.1	Slide Design: Principles	62
5.2.2	Common Supporting Slides	63
5.2.3	Rehearse your Talk!	64
5.2.4	Final Details before the Talk	65
5.3	Deliver the Presentation	65
5.3.1	Delivery: It's Your Stage	66
5.3.2	Handling Questions	66
5.3.3	Managing Time Constraints	67
5.3.4	After the Talk	67
5.4	Attend Presentations: How to Benefit	68
5.5	Posters, Discussions, and Pitches	69
5.6	Bonus: Talking to the Media	70
6	Lecture 5: Crafting the Job Market Paper	72
6.1	The Job Market Paper	73
6.2	The Structure & Writing Guidelines	73
6.2.1	Crafting the Abstract and Introduction	74
6.2.2	Writing Guidelines for Subsequent Parts	77
6.3	Writing and Rewriting	81
6.3.1	Writing Style: Language Guidelines	81
6.3.2	Rewriting as a Process	82
6.4	Final Remarks	83

7	Closing Lecture: Research Products in Economics	85
7.1	Before the Product	86
7.1.1	Selecting a Project	86
7.1.2	Rethinking your Project: Mind the “History Trap”	87
7.1.3	General Advice for Junior Researchers	88
7.2	Academic Products: an Overview	88
7.2.1	Start Early, Revise Often	89
7.3	The Short Academic Product: Abstract and Elevator Pitch	90
7.3.1	Three Elements, One Limit	91
7.4	The Long Academic Product: Introduction and Spiel	95
7.4.1	Canonical Templates and Essential Elements	95
7.4.2	Bonus: Rest of the Paper	98
7.5	Policy-Oriented Research Products	100
	Afterword	103

Preface

I am thrilled to share a few words to support the **Barcelona Essential Skills Training for Economics PhD Students (BESTEPS)**, a brilliant initiative led by an outstanding team of PhD students.

BESTEPS aims to provide a comprehensive set of skills that are essential at various points of doctoral research in Economics. It covers a wide range of topics, from the early stages of the research process to the final production of innovative and relevant research papers. These skills are not typically included in master's or PhD curricula, despite being crucial for a successful academic career and for enhancing the competitiveness of PhD students in the job market and beyond.

In terms of format, BESTEPS is unique in being tailored to the specific needs of Economics PhD students and being highly interactive. The program is structured as a course comprising several sessions, each focusing on a particular aspect of the process of creating, conducting, and publishing academic research. These sessions feature leading researchers sharing their experience. BESTEPS provides a solid foundation for PhD students to approach their research careers in a structured and effective way.

The stars aligned, and this proactive initiative coincided with a funding call launched by the UB Doctoral School to support training activities led by PhD students. The outcome was so successful that we feel compelled to share it. That is why we have put together this *BESTEPS Guide*. We are confident that PhD students from other programs, as well as other junior researchers, will benefit from these notes.

I want to take this opportunity to warmly thank the lecturers, Antonio Cabrales, Dirk Foremny, Albrecht Glitz, Libertad González, Joan Llull, Guillermo Marin, Diego Puga, and Jenifer Ruiz-Valenzuela. They have done a wonderful job sharing their experience in a helpful and enlightening way. They truly exceeded our already high expectations.

I would also like to express my deep gratitude and congratulations to the exceptional team of PhD students who designed the BESTEPS proposal and have meticulously organized every detail of the program with rigor, dedication, and vision. The team, led by Mengwei Lin, includes Daniel Favre, Marcella De

Giovanni, Sergio Blanco, Anna Chesa, Ruixi Zhang, Sofía Balladares, Natalia Carralero, Elisa Muscarella, Thompson Ogajah Tawiah, and many more. Their proactive spirit, commitment to academic excellence, and ability to turn an idea into a high-impact initiative are truly commendable. My sincere thanks also go to Jordi Roca and Judit Farré, whose administrative support has been essential to this program.

Last, to all the young researchers reading this *Guide*: we hope you find it as inspiring and useful as we intended. The research journey can be challenging, but you are not alone; there is a community of scholars ready to support you, and resources like BESTEPS are here to help you grow. Stay curious, stay resilient, and don't hesitate to reach out, ask questions, and seek connections. Your ideas matter, and your work has the potential to shape the future of economic research. Keep going—you've got this!

Concepció Patxot

Director, PhD in Economics Program

University of Barcelona

VOLUME ONE

Set Up Your Research Project

The first part of *The BESTEPS Guide* is designed to support early-career researchers in taking the very first steps of their PhD journey. It focuses on the foundational phase of any research career: setting up a research project. This includes understanding the workflow, finding a promising idea, and developing the mindset and habits with data management and coding practices.

The three lectures in this volume provide complementary perspectives from three experienced economists: Antonio Cabrales, Libertad González, and Joan Llull. In the **Opening Lecture**, Antonio provides a broad overview of the research process, guiding early-career researchers through each stage and offering essential principles for dealing with uncertainty and maintaining focus. In **Lecture 1**, Libertad zooms in on the very beginning of the journey, giving practical, step-by-step guidance on how to generate questions and formulate research ideas. Finally, in **Lecture 2**, Joan focuses on a core competency of today's researchers: doing organized and replicable work, and emphasizing the mindset and practices of reproducibility, transparency, and research integrity.

Volume 1 of *The BESTEPS Guide* focuses on helping you start and structure a research project. Volume 2 will turn to the research product, exploring how to produce meaningful outputs and communicate research effectively.

1 Opening Lecture: Navigating the Research Process in Economics

*Lecturer: **Antonio Cabrales**. Antonio is a Professor of Economics at UC3M, and formerly a Professor and Head of the Department of Economics at UCL. He is also the Executive Vice President of the European Economic Association, a Fellow of the Econometric Society, and a Founding Fellow of the Royal Economic Society. Antonio's research spans a broad range of fields within economics, including game theory and mechanism design, experimental economics, the economics of networks, and the economics of education, among others.*

In this lecture, Professor Cabrales provides a comprehensive framework for navigating the research process in economics, outlining the key stages of a study, highlighting critical points, common challenges, and practical strategies to overcome obstacles. He also shares what he would have liked to know when he started this journey.

Contents of this lecture:

- Introduction: Research as a Power Law
- Step 1: Idea Generation and Evolution
- Step 2: Choosing the Right Methodology
- Step 3: Writing and Communicating
- Challenges along the Way
- Bonus: Preparing for the Job Market

1.1 Introduction: Research as a Power Law

Research follows a power law distribution: there are far more failures than successes. The likelihood of finding the perfect research idea early in your journey is low. Instead of aiming for immediate breakthroughs, you should embrace the

iterative nature of the process. Anticipate challenges and surround yourself with people who can help you navigate them. This is why focusing on creating a safe space and a good work-life balance at the beginning of the PhD will make your experience way more fulfilling.

The following parts of this lecture will cover suggestions for idea generation and evolution, choosing the right methodology, writing and communicating, managing challenges along the way, and preparing for the Job Market.

1.2 Step 1: Idea Generation and Evolution

The first step in doing research is to come up with a research idea¹. A key strategy for this is to maximize your exposure to diverse topics and perspectives in order to seek inspiration. One useful approach is to attend as many seminars as possible, even those outside your immediate field of interest. The value of knowledge is often unpredictable, so what may seem unrelated today might become crucial in the future.

Ideas emerge from multiple sources: reading both classic and recent academic literature, attending conferences, engaging in discussions with peers, seeking guidance from advisors, conversing with scholars from different fields, and even casual talks with family or following news and podcasts. The key is to explore widely, without the need for in-depth analysis at this stage. There is a nice way of doing this in a more “mechanical” way, that follows two steps: *first, think divergently; then, think convergently*.

Divergent Thinking: Generating Ideas

Creativity thrives when there are no immediate constraints. To enhance your ability to generate research ideas, practice brainstorming without self-criticism. The critical mind should not interfere with the generative process too early.

¹The next lecture covers this step extensively.

One useful technique is to write down subtopics within a broad research area. This exercise can help you identify which themes are most promising or personally engaging. The more ideas you produce, the better your chances of refining one into a viable research project.

Exercise 1. Practicing divergent thinking

Since generating ideas is usually challenging at this stage of your career, here are some quick exercises you can perform, either individually or in pairs, to develop your ability to generate research questions:

1. **Brainstorming.** As a starting point, brainstorm individually by writing down all research questions you can think of within 2 minutes, without judging their quality. Then pair up, read each other's questions, and try to generate new ones from them within another 2 minutes.
2. **Mind mapping.** First, pick a broad topic in economics (e.g., climate, gender, immigration, education, taxation, inflation, matching, networks), preferably one you are *not* currently working on, and then map out all possible subtopics and angles within this topic. Doing so will help you discover new and potentially interesting research ideas.

Convergent Thinking: Refining Research Questions

Once you have generated a list of questions, it's time to evaluate them critically. When assessing a research question, imagine a colleague asking: *Why should I care?* Then you can consider some key questions, including:

- Is there a gap in the literature?
- Does the topic address an important social problem or policy issue?
- Can the research contribute to resolving a theoretical debate?
- Is the insight from this study applicable beyond its immediate context?...

Admittedly, not all PhD students have complete freedom in choosing their research topics. However, even if you are assigned a topic from the beginning, you

can still approach it from an innovative angle. Flexibility in framing your research question is essential for making a meaningful contribution. Moreover, you need to find an angle that motivates you, as you will likely spend a considerable amount of time working on it.

Exercise 2. Practicing the Elevator Pitch

Once you have a good idea, you need to both understand it and be able to explain it to others. This step is crucial, since at this stage you need to get as much feedback as possible. A strong research idea should be explainable in 30 seconds. You should prepare a concise “elevator pitch”. To craft your pitch, develop a brief summary answering the following questions:

- Who am I?
- What do I do?
- How do I do it?
- Why does it matter?

This pitch should be accessible to non-experts and should highlight the novelty and importance of your work. A good practice is to try to make someone among your family and friends understand it.

Research Evolution and Sunk Cost Fallacy

You may think you have a great idea and you were also able to turn it into an “elevator pitch”. Yet, research ideas are fluid and will likely change over time. Be open to iteration and refinement, it is part of the process! At the same time, be mindful of the sunk cost fallacy. If a project has been stuck for over a month with little progress, or if you do not anticipate any progress soon, reconsider whether it is worth continuing.

Keeping a steady flow of new ideas prevents stagnation. Therefore, you should continuously alternate between divergent and convergent thinking processes, placing special emphasis on divergent thinking. Then, once one idea seems likely to

fail and a better outside option arises, do not hesitate to pivot. *It is crucial to maintain a pipeline of fresh ideas rather than clinging to a single project at all costs.*

1.3 Step 2: Choosing the Right Methodology

Once you have identified a concrete research question, the next step is to select an appropriate method to answer it. The gold standard in empirical research is the randomized controlled trial (RCT), but this is not always feasible (particularly at early stages of a research career). The optimal method will depend heavily on your specific question (and data availability, if empirical analysis is involved). To answer a question empirically, you might be using:

- Field experiment/RCT (maybe at a lower scale).
- Natural experiment/Instrumental Variables (IV).
- Lab experiments (including online and survey-based experiments, or lab in the field).
- Web scraping
- Administrative data

If you decide to take theoretical approaches, you will have

- Game theory
- Mechanism design
- Matching/coalitional analysis/networks
- General equilibrium
- Computational methods

These lists are non-exhaustive. The guiding principle should be to select a method that is as complex as necessary, *but no more*. Over-complicating a study to showcase technical skills can be counterproductive if it does not align with the

research objectives. Nevertheless, exploring new techniques at this stage can be valuable and signal that you are able to work with more than one.

Exercise 3. Designing your methodology

You are interested in understanding “how variations in climate in recent years are affecting and will affect violence in schools in the future”. Now take 5 minutes to try to design a study to answer that question. It can be theoretical, empirical, or a mix. It can be a general or partial equilibrium.

1.4 Step 3: Writing and Communicating

In the end, as PhD students, you are expected to produce research papers. Even at the early stage, once your idea and execution plan are ready, you need to start writing - otherwise, ideas stay in your mind or in an “Idea # 34543” folder. Effective writing is essential for academic success, though it can sometimes be quite tough. Some useful resources include:

- *The Little Book of Research Writing* (Varanya Chaubey).
- *A Guide for the Young Economist* (William Thomson).
- Papers by well-regarded writers in your field (e.g., Drew Fudenberg)

Clarity is key. Your audience may not be intimately familiar with your topic, so clearly explain why your research matters in an accessible and engaging way. While AI tools can assist with drafting, always refine the output to avoid overly abstract or convoluted phrasing. If your writing seems confusing to you, it will be even more to people who have not spent so many hours on your topic. Remember, you want to reach as many people as possible, so give writing an important slot in your agenda!

Exercise 4. Writing a clear abstract

Here are two versions of an abstract, a clear one and a not-so-clear AI-created one. Compare and discuss.

Version 1: We examine experimentally the impact of communication on trust and cooperation. Our design admits observation of promises, lies, and beliefs. The evidence is consistent with people striving to live up to others' expectations so as to avoid guilt, as can be modeled using psychological game theory. When players exhibit such guilt aversion, communication may influence motivation and behavior by influencing beliefs about beliefs. Promises may enhance trustworthy behavior, which is what we observe. We argue that guilt aversion may be relevant for understanding strategic interaction in a variety of settings, and that it may shed light on the role of language, discussions, agreements, and social norms in these contexts.

Version 2: This study engages with the experimental investigation of the implications that various forms of communication may have on cooperative behaviors and trust-related dynamics. Through a methodological framework that allows for the discernment of commitments, misrepresentations, and cognitive expectations, we analyze the extent to which individual actions align with hypothesized psychological motivations. The findings are interpretable within a framework where actors exhibit tendencies consistent with the minimization of anticipated emotional disutility, specifically in relation to internalized social expectations, a concept that can be linked to psychological game-theoretic modeling. Under conditions where such behavioral patterns are present, communication appears to exert an influence on decision-making processes by shaping second-order beliefs regarding strategic interaction. Notably, the presence of verbal commitments appears to correspond with an increase in behaviors that could be categorized as trustworthy. The broader theoretical relevance of guilt aversion is discussed in relation to its potential explanatory power for a range of strategic contexts, encompassing considerations of discourse, informal agreements, normative structures, and linguistic conventions.

Hints: Pay attention to how the second version over-complicates sentences, adds vague or overly abstract phrasing, and weakens key points.

1.5 Challenges along the Way

Doing a PhD is tough - nobody should tell you otherwise. However, a positive aspect is that everyone around you is experiencing similar challenges, creating a great opportunity to share thoughts and support each other. Here are some common challenges and tips to help you overcome them.

1. Changing directions

Hitting roadblocks is part of the research process. Have you recently encountered a challenge that forced you to shift focus? If so, reflect on what led to the change and what you learned from the experience. Flexibility is a crucial skill for a researcher, and self-knowledge is something you will be learning along the way.

2. Time management

Time allocation is a common pitfall for PhD students. We often misjudge how much time should be spent on different activities. Rationalize your schedule by prioritizing research and carefully managing external demands (meetings, calls, emails, etc.). The administrative burden is something PhD students often underestimate when starting their studies. Keep in mind that substantial time will likely be devoted to tasks such as applying for grants, managing conference budgets, or invigilating exams. Having a good organizing strategy is crucial, and remember that you can always ask senior PhD students how they handle these challenges. Some examples of productivity strategies include:

- **Dedicated Research Blocks:** Aim for two or three uninterrupted 90-minute research sessions every day - no chats, no smartphone, no distractions, nothing else! You can adapt the duration and frequency according to your schedule and maximum attention span.
- **Identify Your Chronotype:** Align your schedule with your natural energy levels, and plan deep work during peak productivity hours.
- **Learn to Negotiate:** Many academics struggle with negotiation. *Getting to Yes* by Roger Fisher and William Ury is a helpful resource. Effective

negotiation skills can help you set appropriate meeting times or manage shared office spaces efficiently.

Exercise 5. Negotiating and “Managing Others”

Scenario: Your office-mate spends large amounts of time on Zoom talking very loudly. It is very distracting. You do not have the luxury of a nice room at home to work, and you need to be at the office. Now get in pairs, and negotiate about the issue. You have 5 minutes.

3. Dealing with Bullying and Harassment

Unfortunately, bullying and harassment occur in academia. If you encounter such behavior, it is important to speak out. Here are some actions that you can take:

1. Confirm that what you are experiencing is bullying.
2. Seek support - most universities have mechanisms in place.
3. Consider both formal and informal complaint routes.
4. Know what to expect after filing a complaint, including potential outcomes, and prepare yourself accordingly.

The most important thing is to feel safe in your workspace during your PhD (and in all aspects of life), so take this as seriously as it deserves. Find someone you trust to talk to, and act promptly!

1.6 Bonus: Preparing for the Job Market - It is too early, and it is not too early

The job market is that scary thing that happens when you are about to finish your PhD. Being ready for it is important, and there are some things that must be considered since the beginning, even though it seems far away. This is because success on the job market requires continuous intellectual engagement, so here are some suggestions that may be interesting to bear in mind since the beginning of your PhD:

- Maintain an idea notebook to capture spontaneous research thoughts.
- Set a weekly goal to generate new research questions.
- Always be prepared to answer: Why is this important?
- Network widely and be open to adjusting your research focus based on feedback.

Overall, the journey you are starting may feel frustrating at times, but it is also one of the most rewarding experiences you can have. By embracing experimentation, critical thinking, and adaptability, you can navigate the PhD effectively and lay the foundation for a fulfilling research career. Two final suggestions would be: 1) Try to concentrate on the activity, not the results. 2) Be ready to explore new ideas and build coalitions.

This was a broad overview of the entire process, so at this point, you may feel a bit overwhelmed - but don't worry! The next lectures will zoom in on specific stages of the journey and help you clarify any doubts about this exciting experience.

2 Lecture 1: Starting a Feasible and Promising Project

*Lecturer: **Libertad González**. Libertad is a Professor of Economics at UPF and BSE, and a research fellow at IZA, CReAM, and CESifo. Previously, she was the President of the European Society for Population Economics and a holder of an ERC Consolidator Grant. Her research spans labor and public economics, focusing on migration, fertility, gender inequalities, and child health and development.*

In this lecture, Professor González focuses on the step of idea generation. As you already know, this is a key step of the process - without an idea, there is essentially no project to pursue. Idea generation is often one of the most anxiety-inducing stages, so getting concrete advice on how to handle it can be particularly helpful! Professor González presents a 6-step process that she thinks is especially effective for developing successful ideas that can become promising papers afterwards.

Contents of this lecture:

- Step 1: Select a Field
- Step 2: Select a Topic
- Step 3: Find and Evaluate a Research Question
- Step 4: Develop a Plan - Approach and Literature
- Step 5: Seek Feedback and Refine the Idea
- Step 6: GO! Move Forward with Your Project
- Bonus: Do's and Don'ts for Junior Researchers

To begin with, we need to overcome the tendency to wait for the “*big idea*”. A common mistake among young researchers is to delay starting a project while reading broadly across topics in search of a perfect idea. Instead, it is essential to

start working on something - feasibility and significance often only become clear after some time and effort. Even if the initial idea does not ultimately work out, the process will deepen your understanding of the literature, methods, and data, making the effort worthwhile!

2.1 Step 1: Select a Field (or two)

The first step in the idea-generation process is to choose a research field, as well as an advisor. This decision is typically shaped by a combination of factors, including your background and training, personal interests, the expertise of faculty members, and the resources available in your department, among others.

Note that your research does not have to align exactly with your advisor's work - you have the freedom to explore your own interests within your chosen fields. In many cases, the fields may already be partially defined when you enter the PhD program, or they may take shape during your coursework.

It is essential to be aware of the various possible fields and ensure the chosen one genuinely motivates you. Remember, *a field is broader than a topic*: for example, labor economics is a field, while gender or migration are specific topics that can be explored within various fields, including labor economics.

2.2 Step 2: Select a Topic

Within your field of interest, identifying specific topics that intrigue you is crucial. Since research involves a long-term commitment, it is essential to select a topic that excites and motivates you, just as it was when choosing your field. Once again, consider your personal interests, the connections between these interests and your field, and the topics covered in your field courses, for instance. Importantly, the topics you choose do not need to be traditionally associated with your field. You can be innovative here!

Exercise 1: List three topics that interest you

Hints: These topics do not have to be “classic” economics topics. They can be

migration, fertility, gender inequalities in economic outcomes, child health and development, poverty, just to name a few examples.

Some more examples: Tax evasion, tax incidence, corporate income tax and automation; populism, social norms, identity and beliefs; climate change, inequalities, subjective well-being.

2.3 Step 3: Find and Evaluate a Research Question

Once you have a topic, the next step is to start formulating research questions within that topic. Remember: a topic is not a question! You can always start out broad from a topic and progressively narrow it down until you arrive at a concrete, *researchable* question.

Here is an example. Say, you define a topic of interest - *migration*. Next, you narrow it down a little bit to, for example, *the economic impacts of migration*, which is still a topic. If you then focus on a specific hypothesized relationship or outcome, you may arrive at a research question such as: *What is the effect of immigration on the wages of native workers?*

Here, for young economics researchers, a practical *tip* is to avoid overly broad “why” questions. Instead, try to formulate more focused questions around specific mechanisms or relationships (Do you have any hypotheses in mind?).

Second, once you have a research question, you need to evaluate it. That is, you need clear criteria to judge whether your question is worth pursuing. Typically, a good research question should be:

- **Relevant:** People care about the question, or you can persuade them to care. At the very least, YOU should care about it.
- **Answerable:** We don’t know the full answer, and so can contribute to closing a gap in the literature.

A common question among early-stage researchers is: Where do research questions come from? One important source is the courses you take. At the PhD level,

taking courses is no longer only about acquiring knowledge. Rather, it should serve as an opportunity for you to come up with your own research questions. Taking courses in different fields may help you combine new insights gained in class and arrive at novel research questions.

If you are no longer taking courses, seminars and conferences become key sources of inspiration. Listening to others present their work can often spark your own ideas. Lastly, non-academic sources, such as news articles, podcasts, or even casual conversations with friends and family, can also be surprisingly fruitful in generating good research questions.

Exercise 2: Write down two research questions based on your chosen topics²

Okay! So now you've completed the first three steps: selecting a field, choosing a topic, and formulating a question. What next?

At this point, it is tempting to begin reading the literature extensively. Indeed, eventually you will need to have a deep understanding of the existing literature in order to identify your own contribution. However, reading too early can be limiting: your thinking may become overly shaped by what has already been done, and this might lead to a contribution that is too incremental.

Instead, you should first think independently about how you would answer your question. That is to say, now is the time to develop a plan. This brings us to the next step.

²You can also revisit Exercise 1 from the Opening Lecture to deepen your understanding of this step.

2.4 Step 4: Develop a Plan - Approach and Literature

1. The Approach: methods and identification strategies

“An idea = a question + the proposed method to answer it.” A well-defined plan that details your method is essential *after* defining your question but *before* touching the data. That said, building such a plan is not always straightforward. A helpful way to get started is by answering the following questions:

- How would you conduct the study with unlimited resources? (What would be the ideal way to answer your question?)
- What skills and methods will be required?

A very useful exercise is to imagine the ideal randomized controlled trial (RCT) that would allow you to answer your question. Try to understand what the randomization units would be, what the treatment would look like, and what outcomes you would explore (This exercise tends to be especially helpful for applied microeconomists but can also be useful for other economists).

This thought experiment will help you visualize the structure of your idea, and, if the ideal RCT is not feasible, find the second-best way to answer the question. For empirical micro research, such second-best approaches include experiments (lab, field, or survey-based experiments) or quasi-experimental methods (Instrumental Variables, Difference-in-Differences, Regression Discontinuity Designs, etc.).

What NOT to Do

1. Don't just find some data and start playing with it and running regressions!
2. Remember: simply exploring the data and conducting exploratory regressions is *not* an identification strategy.

Be ambitious and proactive. If the data you need is not available, look for ways to obtain it or consider generating your own. If you lack certain technical skills, take the initiative to learn them. If the project is time-consuming, start

early. If funding is needed, explore potential sources and seek advice from advisors.

But most importantly, leverage your *unique* perspective: Your background, training, experience, and network provide you with a comparative advantage. Think about research settings and perspectives that you are especially well-positioned to explore.

Key Questions for identifying an approach

1. What would an ideal Randomized Controlled Trial (RCT) look like for answering your question?
2. If an RCT is not feasible, can you identify a natural experiment or quasi-experimental setting?
3. Or, which methods are appropriate (e.g., theoretical models, structural models, causal inference techniques, experiments)?
4. What assumptions will be needed to answer the question credibly?

2. Conducting the Literature Review

Finally! Once you have drafted a research plan, conducting a literature review helps you understand what others have done and identify your potential contribution. You should expect to find papers addressing similar questions - this is entirely normal (after all, many people are doing research!). The key is to identify what you can do differently or better.

However, bear in mind that you should avoid doing the following (as they do not constitute a meaningful contribution):

- Simply replicating a study using data for a different country, without a clear motivation for what new insights could be obtained.
- Claiming to “add to the literature on something” without a clear innovation.
- Writing a model (per se) without empirical relevance or support.

- Claiming to be the first to study something.
- Reporting mixed results in the current literature without a compelling explanation of how your project improves on it.

A true contribution should provide new knowledge that was previously unknown, improve existing methodologies, or offer more credible identification.

2.5 Step 5: Seek Feedback and Refine the Idea

Okay, now you have a clear question and a structured plan to answer it. However, before moving on to data analysis, it is important to discuss your idea with supervisor(s), peers, and other faculty members. Their input can help you refine both your research question and your approach.

One of the key roles of your *supervisor(s)* is to leverage their experience and expertise to help you evaluate whether an idea is worth pursuing. So listen carefully to their suggestions! (Fortunately, at UB we also have our *Research Ideas Coffee Session Series*, which provides an excellent opportunity to receive feedback from your peers. Make sure you participate actively and benefit from it!)

It is worth repeating that this step is crucial. Keeping your ideas to yourself might cause you to miss valuable perspectives or overlook potential issues. Feedback is an iterative process and may lead you to revisit earlier steps. That's not a waste of time - on the contrary, it's part of becoming an expert in the research process (which is actually what a PhD is for). So don't be afraid to ask for help and comments!

The Feedback Loop

After receiving feedback, you will often find yourself at a crossroads. Your advisor, or other people you consult, may give you one of three types of responses:

- **Great idea!** If the project is well-motivated, feasible, and methodologically sound, go ahead and proceed to execution (proceed to Step 6).

- **The project is not worth pursuing.** Perhaps the contribution is too minor, the identification strategy is not credible, or the question is infeasible. (Go back to Steps 2 or 3, and refine your topic or question.)
- **Unclear.** If they are unsure, talk to a few more people, or go back to Step 4 and work further on your plan. Then repeat Step 5.

Remember: iteration is part of the process. Hearing “no” or “not sure” is not failure - it’s a prompt to think deeper and improve your idea.

2.6 Step 6: GO! Move Forward with Your Project

Arriving at this point is great news. With the “OK” from your advisor, you now have something to work on! While the execution stage is beyond the scope of this lecture, it is worth noting that there will be many back-and-forths, so make sure to ask for feedback frequently. You can present at internal seminars, talk with additional faculty members, visitors, and more senior students in close fields. From here on, the focus shifts to other key aspects of the research process, which will be addressed in the upcoming lectures.

2.7 Bonus: Do’s and Don’ts for Junior Researchers

To wrap up the lecture, here is some extra advice from Professor González. First, a few common practices that may be unhelpful during the PhD, or when pursuing an academic career in general:

- Shying away from discussing your project: Engaging with others helps refine your ideas and improve research quality (we’ve already emphasized how important feedback is!).
- Rushing into data analysis without a clear plan: Always develop a well-defined identification strategy first.
- Starting with a dataset or policy change rather than a research question: Research should be driven by important questions, not just available data (though data and context can certainly inspire good questions).

- Following trends instead of personal interests: Study what you are genuinely passionate about.

And also, some general advice that should benefit your PhD journey:

- Keep your advisor(s) regularly updated.
- Get to know your data REALLY well (read the codebooks, explore descriptive statistics, ...).
- Start from a *relevant* question.
- Work on a topic that you truly care about.
- Think hard about your (potential) contribution to the literature.

The research process is iterative and requires persistence, creativity, and flexibility. Your initial idea will likely evolve over time, and that's perfectly fine. What matters is keeping your research question and approach clear. This will help you develop a solid project and ensure a meaningful contribution to the field. And most importantly, HAVE FUN!

3 Lecture 2: Doing Organized and Replicable Research

*Lecturer: **Joan Llull**. Joan is a Professor of Economics at IAE-CSIC and BSE. He is the Data Editor of *Econometrica*, *Quantitative Economics*, and *Theoretical Economics*. He was awarded an ERC Starting Grant and an ERC Consolidator Grant. His research is mainly in labor economics and structural microeconometrics, focusing on migration, inequality, human capital accumulation, technological progress, and labor market frictions.*

In this lecture, Professor Llull will present the replication package as a conceptual and practical framework for doing organized research through better managing data and coding. He provides an overview of replication packages and goes through concrete steps in building a replication package, which is not just to make it easier for others to reproduce your work, but also to help yourself keep track and build on your own work.

Contents of this lecture:

- [Introduction: Reproducibility and Replication Packages](#)
- [Part 1: Managing Data](#)
- [Part 2: Coding Organized Projects](#)
- [Bonus: Main Takeaways and Ten Simple Rules](#)

3.1 Introduction: Reproducibility and Replication Packages

Organizing the project according to the standard and structure of a replication package is not just about preparing for journal requirements or increasing replicability for others, but also having a disciplined framework for organizing the entire research workflow. Rather than treating it as a final-stage obligation, we should

use the logic of a replication package to guide our work from Day 1: organizing and documenting raw data, scripts, and outputs with the end product in mind³.

Reproducibility and Replicability in Economics

Reproducibility, defined as obtaining the same results using the same data and same methods, is what most journals verify before publication. It is the cornerstone of transparent and credible empirical research, allowing the profession to build incrementally and confidently on prior work. However, reproducibility is just the first step. The broader profession may later explore replicability, a more demanding standard: whether the same results can be recovered using new data, different methods, or both.

The responsibility of ensuring reproducibility falls not only on journals but also on us, the researchers. Reproducibility is directly tied to the impact of your research. Papers with accessible and well-documented code are easier for other researchers to engage with, cite, and extend. If you want your research to really have an impact, you should allow others to reproduce it and build on it.

In economics, a number of initiatives are actively promoting reproducibility and replicability. Journals now often require data and code availability statements, and some have appointed dedicated Data Editors to conduct pre-acceptance reproducibility checks. The Social Science Data Editors have developed shared standards, such as the [DCAS](#) and a template [README](#) file, to harmonize expectations across journals. Public platforms like the [Replication Wiki](#) curate over 16,000 links to published replication packages, while organizations such as the [Institute for Replication \(I4R\)](#) carry out post-publication checks and host Replication Games, bringing students and researchers together to test findings.

Replication Package: a Framework for Organizing a Project

When we refer to “replication packages”, we are actually talking about “reproduction packages”, in the sense that these packages typically allow readers to

³Although it’s not covered in this Lecture, you can also find information on how to track version control of a co-authored project [here](#).

use the same data and follow the same methods to obtain the same results of a paper.

A replication package is not just a technical necessity; it's a matter of integrity and communication. Think of it as a **teaching tool** for your work, offering a clear step-by-step guide to how you reached your findings. Also, remember, future-you will thank you when revisiting the project for revisions or extensions. In contrast, an opaque or confusing package will frustrate users, diminish the impact of your research, and even prevent yourself from understanding the previous exercises.

It is essential to develop the habit of **creating a replication package from Day 1 of a project**. This offers several advantages.

1. It helps your future self: We often need to revise and update, and having the project organized towards a replication package makes this easier.
2. It fosters scientific openness, and you will be requested to publish your package at some point.
3. Packages that pass pre-acceptance checks often speed up the publication process (surprisingly, about 75% of submissions fail at least one check).
4. Replication packages increase the visibility and impact of your research: Data from the Econometric Society shows that a typical package is downloaded around 100 times per year.
5. A well-organized package is less likely to contain unnoticed errors and boosts the credibility of your research.

Structure of the Replication Package

But what exactly should we do to organize the project folder as a replication package? Or, what does a replication package contain? An ideal replication package typically contains **16 standard elements**⁴ according to the Data and Code Availability Standard (DCAS), which aims at harmonizing the requirements

⁴See [DCAS](#) for a full list.

from all journals. Broadly, a replication package (and your project folder) consists of **data, code, and documentation**⁵:

- **Data:** Include a Data Availability Statement, raw datasets, and, where relevant⁶, the processed datasets used for analysis.
- **Code:** Provide all scripts needed to transform raw data into analysis-ready data, and to produce the final tables, figures, and other results.
- **Documentation:** This ties everything together. It should explain how to run the scripts, how to access external data (when it cannot be shared), and how any experiments or surveys were conducted, etc. At a minimum, include a **README file**.

Next, we will review the **best practices** for creating a good replication package (and for organizing your project), divided into two parts. First, we'll look at how to manage your data. Second, we'll walk through how to organize your project files on your computer. Each part is presented as a step-by-step guide.

3.2 Part 1: Managing Data

At its core, data management is about enabling replication. Even with flawless code, replication is impossible without the underlying data. Therefore, the first part of this lecture is devoted to how to manage data, divided into three steps: knowing the principles, dealing with different types of data, and citing and enabling access.

Step 1: Know the Principles

The guiding rule is simple: **Make your data available**. In case this is not possible due to legal restrictions, cite the data properly and describe how others can access it.

1. General Principle

⁵See Section 3.3 for a detailed recommended folder structure in practice

⁶Meaning when creating the analysis-ready data is a lengthy process.

- As a default, include all datasets necessary to replicate your results.
- Exceptions apply only when you lack permission to redistribute the data, or when your exact data extract is already available in a trusted repository (see below).

2. **Best Practices for Sharing Data:** To ensure others can fully understand and reproduce your work, share:

- Copies of your datasets in non-proprietary formats (e.g., `.csv`, `.dta`, `.txt`, `.parquet`).
- Raw data, along with cleaning code and documentation.
- Comprehensive metadata and data dictionaries describing variable names, units, coding schemes, and data sources. This is especially crucial for datasets with abbreviations, survey codes, or administrative data.

If you cannot share certain datasets, explicitly state this in both the README and metadata files, and explain the access conditions (e.g., through a formal request process).

3. **Trusted Repositories and the FAIR Principles:** If you don't upload the data yourself but it already exists in a repository, the repository should adhere to the FAIR principles:

- **Findable:** Searchable via metadata and a persistent identifier (e.g., DOI).
- **Accessible:** Downloadable using standard open protocols.
- **Interoperable:** Compatible across different software environments and integrable with other datasets.
- **Reusable:** Distributed under a clear license, with documented provenance and in standard formats.

Trusted repositories include Zenodo, Dataverse, and OpenICPSR. Platforms such as GitHub, Dropbox, or personal websites do NOT meet FAIR standards.

Step 2: Dealing with Different Types of Data

1. **“Raw Data”**: Understanding the meaning of raw data is essential in empirical work:
 - Raw data **includes**: Datasets saved in a different format or with unused variables removed.
 - Raw data does **not include**: Datasets where observations were dropped or variables were transformed.

This matters because raw data should allow others to recreate your cleaned dataset exactly from the original extract. The best practice in managing raw data is to keep it in read-only mode to prevent accidental changes. Besides, for historical or scanned data sources (e.g., old PDFs or digitized books), include them in a dedicated **Documentation** folder.

2. **Non-Redistributable Data**: If you cannot share certain data, you must request an exemption — typically during journal submission. Common reasons include using confidential administrative data, commercial restrictions, or public datasets with non-redistribution clauses. Even in these cases:
 - Others should be able to apply for access under the same terms.
 - The data extract should be archived (e.g., institutionally) to preserve reproducibility.
 - Ideally, there should be no public alternative data that could yield equivalent insights.

A good practice when sharing real data isn’t feasible is to provide **synthetic or simulated data**:

- This enables users to run the code and understand the workflow, even if they cannot replicate the exact results.
- Particularly useful for complex data structures such as panels, spatial data, or matched samples.

3. **Self-Collected Data:** If the data is collected by yourself, you cannot restrict it for your exclusive use after publication. You must:

- Share relevant materials (e.g., questionnaires, methodologies).
- Archive the raw, anonymized survey data (rather than just the cleaned version).
- Follow IRB guidelines for anonymization and data publication.

Step 3: Citing Data and Enabling Access

A well-crafted Data Availability Statement (DAS) in the README file (more details about the file will come later) is essential. It provides transparency about every dataset used, including those you do not include in the replication package. Your DAS should cover:

- **Provenance:** The origin of the data — who collected it, when, and for what purpose.
- **Access information:** Where the data is hosted, with working links when possible.
- **Instructions:** If the data is not directly downloadable, explain how others can obtain it (e.g., by submitting a form, emailing the provider, paying a fee, or applying for access).
- **Primary data:** If you collected the data, describe your data collection procedures in detail, including sampling design, timing, questions, and instruments.
- **No data sharing?** Then your DAS must be especially detailed, making it clear what data is unavailable and why.
- **Clarity:** Ensure that readers can easily match datasets listed in your DAS with the files provided in your replication package.
- **Citations:** Include a references section in your README file, just as you would in a paper, to properly cite data sources and providers.

Ultimately, data sharing is fundamentally about **trust and transparency**. It demonstrates to the profession that your findings are not the result of a black box. By sharing the pieces (data, code, and documentation) you allow others to reconstruct the machine and confirm that it works. It also means thinking beyond immediate publication, toward the long-term reproducibility and accessibility of your work.

3.3 Part 2: Coding Organized Projects

As discussed in the Introduction, the replication package can serve as the central framework for organizing a research project — not something assembled at the end, but a structure that takes shape from Day 1 and evolves as the project develops. Delaying the creation of replication materials until after a paper is conditionally accepted often leads to disorganization and increases the risk of errors.

Step 1: Set up an Organized Project Folder

Imagine a future replicator (a reviewer, a PhD student, or even your future self) who encounters your code without prior context. Will they understand your logic, structure, and reasoning? Would they be able to replicate your results confidently?

This mindset of **computational empathy** encourages you to design your replication package with clarity and user-friendliness in mind. Every decision you make, from file organization to code comments, should aim to reduce the effort required for someone else to reproduce your findings.

In this spirit, establishing a clear and consistent folder structure lays the foundation of a replication package and for your own efficient research. It helps reduce the risk of losing track of steps, simplifies the revision process, and fosters good habits from the beginning. Here's an example of a recommended folder structure for a Stata user:

```
Project_Folder (Future Replication_Package)/
```

```

1_Data/
  Raw/          # Downloaded, unaltered data (e.g. CPI indices, shapefiles
                , surveys)
  Cleaned/      # Final datasets used for analysis (often left empty for
                replication uploads)

2_Code/
  2025_01/      # Versioned by month; create a new folder when starting a
                new round of work
    Do.files/   # Main scripts organized by purpose
      0_Main.do
      1_Clean.do
      1_CL_undoc.do # A special cleaning step, if applicable
      2_Descriptives.do
      2_DS_cohort.do
    Log.files/  # Logs for each run
    Tables/     # Tables generated during analysis
    Figures/    # Figures generated during analysis
    Stata_packages/ # Custom Stata ado directory for package
                    reproducibility

  2025_02/
  ...

3_Documentation/
  README.txt    # Instructions and metadata for replicators

4_Output/      # Optional: If tables and figures are too large, store
                final outputs here
                # Otherwise, keep them within the corresponding Code
                subfolders

```

Here, separating raw data from cleaned data prevents confusion and accidental overwriting. Naming code folders by year and month (e.g., 2025_01) helps track

the evolution of the project. When moving forward, simply copy the code folder from the previous month to make a new one for the following month.

Remember, all we do here to set up an organized project folder from the very beginning isn't just about replication; it's about working efficiently and transparently from the very beginning.

Step 2: Code in an Organized Way

Zooming in on the **Code** folder, computational empathy is again central here. You should include all the code for data transformation and analysis. This makes explicit all steps and decisions, and allows readers to inspect your methods and reuse them more easily.

1. Programs to include in the folder

- Data cleaning scripts that process raw files into usable formats.
- Scripts for merging or reshaping datasets, and for constructing derived variables or indices.
- Estimation and model-solving codes.
- Simulation and visualization codes.
- Code that generates tables, figures, and supplementary appendices.

If you use manual software such as ArcGIS, you should also provide detailed step-by-step instructions for your exercises. However, it is recommended to use scripts whenever feasible.

2. Have a master file

Moreover, the **Code** folder is not just a collection of scripts, but a complete, transparent record of your computational workflow. To achieve this, begin by having a “**master file**” that runs everything, typically named `0_Main.do` (or `.jl`, `.py`, `.m`, `.R`, etc.). Ideally, this script is the one that:

- Sets up the computational environment by defining paths, loading required packages, and setting global macros or variables.

- Sequentially calls all other scripts in the correct order.
- Automates the entire analysis pipeline, from processing raw data to generating final tables and figures, so that a single run can reproduce all your results.

3. Pay attention to paths

Always use relative paths. Set global paths at the top of main scripts (for example, in Stata: `global path "C:/MyPath/ReplicationPackage"`), and avoid absolute paths that tie your package to a specific computer, and write relative paths like `"$path/Data/Raw/census.csv"` so others can easily re-run your code by changing the first main line in the master code that defines the path. When writing paths, use **forward slashes (/)** to enhance cross-platform compatibility.

Bonus Tips: Extra good practices that show computational empathy

- Use subfolders generously to organize code, data, and outputs neatly.
- Keep individual scripts short and split long ones into sub-files.
- Prefer open, non-proprietary formats like `.csv` or `.xlsx` for data files.
- Keep track of dataset versions, dates, and download links.
- Give meaningful names to programs, variables, files, outputs, etc: use `avg_age`, not `v22`.

Step 3: Keep Track as the Project Moves On

As your project evolves, it's crucial to maintain clear, organized, and reproducible records of every step. This means not only writing clean, well-commented code but also ensuring that others (and your future self) can understand, replicate, and trust your entire workflow. Remember, computational empathy!

1. Document every step to run the code

Every script you write along the way should be thoroughly documented.

Consistently comment (and indent) your code to explain the logic and structures.

In particular, build the **README file** as you go to explain how to reproduce your work. Basically, it should include:

- The Data Availability Statement.
- A list of all datasets (included and omitted).
- Computational requirements (e.g., software versions).
- Descriptions of each script or program.
- Instructions for replicators (including run order).
- Citations for any reused data or code.

Following the Social Science Data Editors' [README template](#) can save time and help prevent errors. With the README file, you are able to provide a high-level overview of the program structure, list manual actions, and link outputs to scripts and files systematically. You should also cite third-party libraries and data in the README.

2. Describe software and hardware requirements

Even perfect code can fail replication if others don't know what environment is needed to run it. Anticipate these questions:

- Does the code run on a standard laptop, or does it require a workstation?
- How long will it take to run? Seconds, minutes, hours, days?
- What software and packages are needed (and their versions)?

To make these clear for the replicator, in the README file, you should:

- Specify precise software versions (e.g., Stata 17 SE, R 4.3.1, Python 3.10 with pandas 2.1.0).
- List all critical libraries and packages - if the code breaks without them, name them; document their versions, as well as installation instructions.

- List hardware used (e.g., Mac M2 with 16GB RAM) and note if it can run on less powerful ones.
- Estimate running time by code section (e.g., data cleaning: 2 hours; regressions: 10 minutes).
- If your package is too time-consuming: prepare a package that allows for a partial replication (e.g., include already-clean data in an intermediate file).

3. Automate and save everything

Automation not only ensures consistency but also saves time when revisiting or updating the analyses. Manual steps, like copying regression outputs into tables or figures, introduce errors and reduce reproducibility. Instead:

- Automate the generation of all tables and figures directly from code.
- Save all outputs automatically and compute all in-text results from code.
- Record logs and diagnostics for every script, and include them in your replication package in an accessible format.
- Automate package installations if possible.

In summary, you need to keep track of your own practices and provide clear instructions, with the goal of guiding the replicator in your mind; don't assume that users will "figure it out".

Step 4: Wrap Up with the Replication Package

Upon completion of the project (especially for a journal submission), it's time to finalize and polish your replication package. This step goes beyond simply re-running your code: it's about ensuring that your entire workflow is clear, reproducible, and ready to be shared with the world.

1. List all exhibits that can be reproduced in the README

Transparency requires spelling out exactly what outputs your package will reproduce. Include a table that maps exhibits (e.g., tables and figures) to the

corresponding scripts, output files, and line numbers. Indicate whether each exhibit requires confidential data. This gives replicators a clear roadmap of what to expect.

Table 1: List of Reproducible Exhibits

Exhibit	Script	Output File	Line Number	Confidential Data Needed?
Table 1	<code>regressions.do</code>	<code>table1.tex</code>	line 45	No
Figure 3	<code>plots.jl</code>	<code>figure3.pdf</code>	line 101	Yes

2. Include all supporting materials

Now, before finishing your package, double-check whether it includes all the supporting materials. You should (whenever applicable):

- Include survey instruments or experimental instructions
- Original questionnaires, translations, sampling schemes, and reward structures.
- If you used software (Qualtrics, SurveyCTO, zTree), share configuration files.
- Include ethics approvals (IRB protocol number, institution).
- Share pre-registration links (permanent URL or DOI).

Remember, you should not include the manuscript itself unless journal policies explicitly allow it.

At this point, also double-check if you have the complete and up-to-date README file. Then make sure that you use standard formats (PDF, TXT, etc.) for documentation files where possible.

3. Tag a permissible license

Remember, sharing the replication package doesn't mean readers can use it without permission. A license clarifies what others can do with your code and data:

- **For data:** Use a Creative Commons license (e.g., CC-BY for reuse with attribution, or CC0 for public domain).

- **For code:** Use standard open-source licenses such as MIT or BSD.

You should document the license clearly in your README file. Clear licensing not only boosts reuse, citation, and impact but also reduces legal ambiguity. At a minimum, ensure that your license allows replication, but check repository or journal policies—some impose specific licenses.

4. **Finally, re-run everything once again**

Once you believe everything is in place, before zipping and sending it, perform a full dry run of your replication package. This process often reveals hidden issues such as hard-coded paths, undocumented steps, or missing dependencies. The goal is to confirm that your package runs seamlessly from start to finish, without any manual intervention. This can help you detect simple mistakes such as typos, file misnaming, and version mismatches.

- Run the main script from top to bottom.
- Ask a non-involved colleague or RA to test the package without your guidance - this mimics the true experience of an external replicator.
- Run it in a clean environment (different machine if possible).
- Confirm that everything runs from the package only.

This final check helps you have fewer and faster iterations with Data Editors and leads to a faster publication process.

If your research process has been organized from the start, the replication package basically builds itself - just zip the project folder, and you're done! Magic! Now, the last thing to do is archive your replication package in a reliable, long-term, and citable repository. Choose a repository that issues a DOI, making your replication package easy to cite. If the journal does not specify a platform, you can consider OSF (Open Science Framework), Dataverse, Zenodo, and AEJ Data and Code Repository (for AEA journals). Remember, you are the one responsible for the package uploaded there, not the journal!

3.4 Bonus: Main Takeaways and Ten Simple Rules

Over the past years, we have seen a solid trend for Economics journals to ensure reproducibility, which is a necessary step to encourage replication. This lecture has explained the importance of reproducibility and how to build a replication package for an economics research project. It is worth repeating by the end of the lecture that documentation (README file) is essential. Also, keep the raw data, carefully read the terms of use, and, if necessary, apply for exemptions at first submission.

Let's now conclude this lecture with the "10 simple rules for creating a replication package" (Vilhuber, Koren, Csóka, Connolly, and Llull):

- Rule 1: Computational Empathy!
- Rule 2: Make your data available.
- Rule 3: Cite data and describe how others can access.
- Rule 4: Describe software and hardware requirements.
- Rule 5: Provide code for data transformation & analysis.
- Rule 6: Explain how to reproduce your work.
- Rule 7: List all exhibits that can be reproduced.
- Rule 8: Include all supporting materials
- Rule 9: Tag a permissible license.
- Rule 10: Re-run everything once again.

In the end, a complete and well-documented replication package is more than just a part of the submission. As researchers, following the steps to build your replication package the day you start your project creates a perfect opportunity to ensure an organized research process. In this sense, a replication package serves as a framework for organizing a research project!

VOLUME TWO

Produce and Communicate Your Research

The second volume of *The BESTEPS Guide* is designed to support early-career researchers in taking the very first steps of their PhD journey. **Volume 2** of *The BESTEPS Guide* focuses on turning research ideas into clear, credible, and well-communicated research outputs.

Building on the foundations laid in Volume 1, this volume addresses the skills required to communicate and produce research outputs: visualization, presentation, and writing. In **Lecture 3**, Guillermo Marin and Dirk Foremny discuss how visualization can be used both to think about data and to communicate economic results effectively, moving from general design principles to applied examples in Economics. In **Lecture 4**, Jenifer Ruiz-Valenzuela examines how to present research effectively, covering slide design, rehearsal, and delivery of a talk, as well as other formats of economics presentation. In **Lecture 5**, Albrecht Glitz focuses on how to write a strong Job Market Paper, explaining the role of a JMP and providing writing guidelines. The **Closing Lecture** by Diego Puga reflects on various formats for delivering our research products, providing insights into project selection, evolution, and communication formats.

The two volumes of *The BESTEPS Guide* together aim to map out the whole research life-cycle during the PhD journey and provide practical guidance on each critical stage of a research project.

4 Lecture 3: Visualizing Economic Research

*Lecturers: **Guillermo Marin** & **Dirk Foremny**.*

Guillermo is the Research Line Lead of “Data Visualization and Storytelling” at the Barcelona Supercomputing Center (BSC). Guillermo specializes in transforming complex datasets into informative and compelling narratives across scientific and public-facing contexts. He brings a complementary perspective for us as economists.

Dirk is an Associate Professor at the University of Barcelona, the Co-director of Tax Systems Analysis at the Barcelona Institute of Economics (IEB), and a Fellow at CESifo. His research focuses on empirical public finance, including topics such as taxation, mobility, and fiscal capacity. He brings us an aesthetic awareness of visualization in economics, along with tools he has found especially useful.

In research, well-designed visuals play an important role in the stage of interacting with data, helping us detect unexpected patterns, test the plausibility of models, reveal heterogeneity, and more. It also allows us to communicate clearly with both technical and non-technical audiences. The lecture focuses on visualization, organized into two parts. In [Part 1](#), Guillermo Marin introduces a general framework of visualization, covering perception, visual encoding, and the design principles. In [Part 2](#), Dirk Foremny builds on this from the perspective of an economist, showing how figures can be used to explore data and present results.

4.1 Part 1. Introduction to Visualization

This part discusses the general framework of data visualization, showing how a common set of techniques applies across fields and offering a conceptual foundation for practice. It also presents the logic of visual encoding—how data is transformed into graphics through various channels, and the perceptual limits that make some designs clearer than others, illustrated with real-world examples.

Contents of Part 1:

- Visualization: Power and Pitfalls
- Understand Visual Encoding
- Best Practices for Academic Visual Design

4.1.1 Visualization: Power and Pitfalls

Why visualization at all? Why don't we just use tables or text? To answer this question, we will start from a definition of visualization: the visual representation of complex information in a way that helps humans to reason and comprehend. From here, we can explore the power and pitfalls of visualization, which hopefully can lead us to an answer in the end.

1) Visual Thinking as a Cognitive Tool

Well-designed graphics augment human intelligence by making structures visible. As an example, check Figure 1: in just a glance, you can decide whether to read the book instead of watching the movie (e.g., Ender's Game, great book but mediocre film) or watch the film instead of reading the book (e.g., the Godfather, outstanding movie even compared to the highly-rated book). This is the power of visualization: it lets us extract useful insights almost instantly, something that would take much longer if we only had a table of scores.

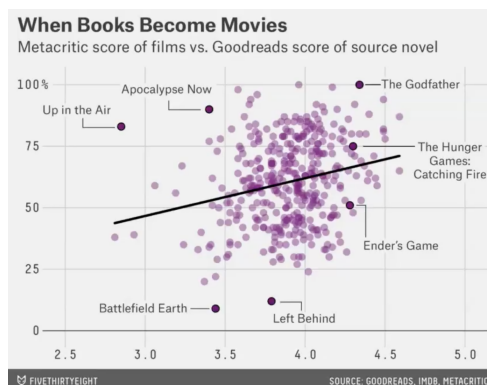


Figure 1: Book vs Movie Scatterplot



Figure 2: A Post on Homicides

2) When Visualization Misleads

Poorly designed graphs can distort or obscure our thinking. A misplaced base-

line, an inconsistent distance, or a truncated axis can radically alter perceptions. A clear example is the post shown in Figure 2, which compares collective homicides in Colombia during 2010–2018 with those in 2019–2020. Presenting deaths over two years versus over eight years without clarifying time scales makes the reduction look much larger than it really is. The same visualization techniques can also produce lies!

Color choices matter. In the example below (see Figure 3), the same dataset is shown with three different encodings: the rainbow palette (left), a single-hue palette (middle), and a gradient of differences (right). The rainbow palette makes it look visually striking but is perceptually misleading, while the more uniform palettes display the true underlying patterns more clearly. Here, since human vision does not perceive equal steps in hue as equal steps in value, differences in green and yellow appear much stronger than differences in red or blue. This can artificially suggest gradients or boundaries that *do not actually exist* in the data, distorting interpretation just as much as truncated axes or inconsistent baselines do.

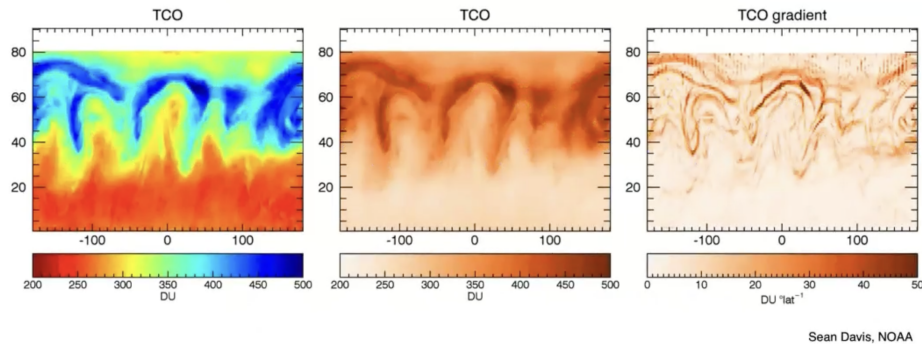


Figure 3: The Rainbow Color Palette Example

Cluttering can lose graphs the power they have. When too much information is packed into a single graph, be it too many lines, overlapping labels, or unreadable axes, the message disappears. Instead of clarifying, the figure overwhelms and confuses the viewer. The examples in Figure 4 illustrate this problem: on the left, multiple series overlap chaotically with an unreadable x-axis; on the right, labels cover each other so densely that the underlying spatial pattern cannot be seen. In both cases, the chart fails to serve its purpose. The lesson is simple:

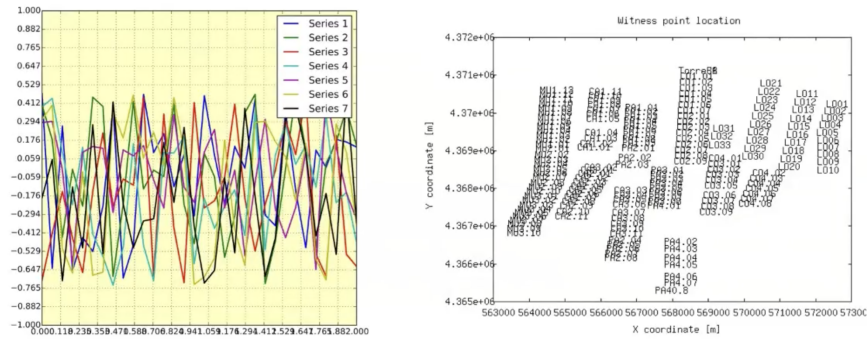


Figure 4: Cluttering Interferes with Legibility

less is more. Good visualization requires prioritizing clarity over exhaustiveness.

3) Use Visualization in Our Favor

Now that we know **visualization can help us think, and that it can also make us think wrong**, we can better use it to our advantage. Figure 5 is an example where a graph is used to help us think easily and correctly. It shows measles cases in the United States by state and year. Each small square represents one data point, with colors reflecting incidence. The vertical black line marks the introduction of the measles vaccine.

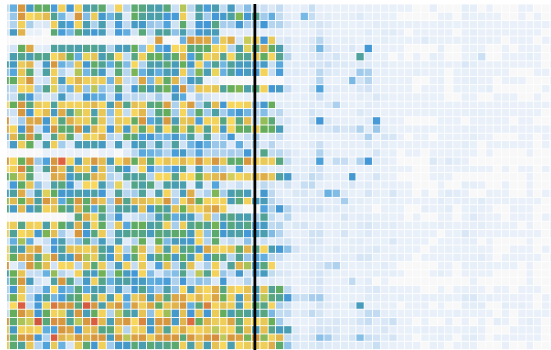


Figure 5: The Impact of Vaccines on Measles

With this design, the viewer grasps at a glance what would be less clear in raw numbers: before the vaccine, measles outbreaks were widespread and frequent. After the vaccine, the disease virtually disappears. This shows the power of visualization when used properly. It helps us to:

- Spot patterns and breaks (e.g., before vs. after a policy).
- Summarize complexity in ways tables cannot.
- Communicate persuasively to diverse audiences.

From this, the answer to why we want to visualize is clear: visualization can increase human abilities to overcome cognitive and memory limitations and help us think more clearly!

4.1.2 Understand Visual Encoding

We have seen that visualization can both clarify and mislead, and that a graph is an **external representation** that extends human cognitive capacity. Now we will further explore visual encoding. This section draws heavily on *Visualization Analysis and Design* by Tamara Munzner.

1) The Visualization Process

Tamara Munzner defines visualization systems as providing visual representations of **datasets** designed to help **people** carry out **tasks** more effectively, and models the visualization process as follows:

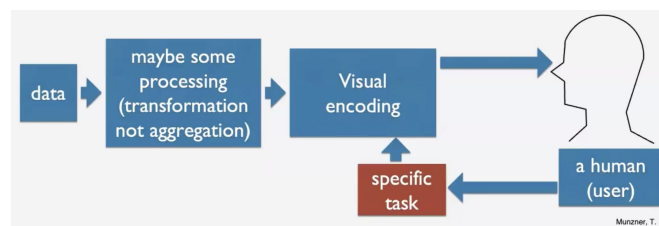


Figure 6: The Visualization Process

Within this framework, visual encoding sits at a critical point, representing the process of mapping data into visual information that helps us humans with specific tasks. But before discussing visual encoding, it is also important to note two other characteristics of the visualization framework:

1. **Human in the loop:** Visualization systems are most appropriate when the goal is to augment, not replace, human reasoning.

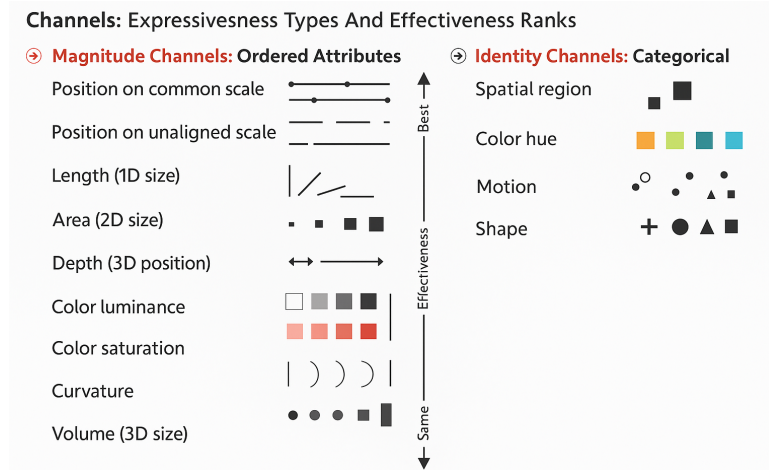


Figure 7: Channels (from *Visualization Analysis and Design*)

2. **Task orientation:** Visualization always serves a purpose: comparing values, detecting outliers, and identifying patterns. The design must match the task. Additionally, when we are unsure of the exact question to ask, visualization helps us explore. Algorithms cannot do this better or faster.

2) Visual Encoding: Channels, Expressiveness, and Precision

Now we turn to visual encoding. It is worth emphasizing that the graphic elements we choose matter! Figure 7 helps us understand how these elements affect the expressiveness of our visual design:

Here, visual channels are separated into magnitude and identity, and the vertical axis highlights the effectiveness of certain visual encoding:

- **Magnitude channels** are for attributes with some inherent order, ordinal or quantitative, showing *how much*. These include position, length, angle, area, depth, luminance, saturation, curvature, and volume. They are ranked vertically by effectiveness in the figure, from most effective (e.g., position) at the top to least effective at the bottom (e.g., volume)
- **Identity channels** are for categorical attributes, providing information about *what*. These include spatial region, color hue, motion, and shape. They do not imply order but are effective at distinguishing categories.

Psychological research confirms that not all encodings are equally precise. According to Stevens' Power Law, our perception of magnitude is nonlinear: we may overestimate some differences and underestimate others. Experiments consistently find that people are most accurate when comparing position along a common scale, less so with length or angle, and least with area or volume.

The lesson from understanding visual encoding is straightforward: **the graphics we choose matter**. When visualizing our data, the chosen channel must match the data attribute: using color hues for quantitative attributes can mislead. Besides, whenever appropriate, choose the elements with higher effectiveness and metrics that human brains can perceive more precisely.

4.1.3 Best Practices for Academic Visual Design

Based on our understanding of visual encoding, we now know that designing effective graphics is not about aesthetic choices alone but about following principles that make information clear, precise, and interpretable. Here are some recommended practices:

1. **Pay Attention to the Axes:** Axes can radically alter how trends are perceived. Always show a complete axis for charts that encode *size* (e.g., bars); they must start at zero. Line charts, by contrast, encode point *position*, so a truncated axis may be acceptable for highlighting variation. Aspect ratios also matter. The slope can look dramatic when the axis is compressed, or negligible when stretched. You may want to use a wider ratio (to emphasize long-run trends) or a narrower one (to highlight short-term spikes). Research shows that perception is more precise at a 45° angle.
2. **Annotate Thoughtfully:** Never assume that your audience knows your data! Always provide clear axis labels, titles, legends, and color scale. Highlight key points when necessary.
3. **Show the Data, Not Just the Statistics:** Summary statistics sometimes are misleading. A famous illustration is the [Anscombe's Quartet](#): identical descriptive statistics, radically different data. Always check the distribution to avoid missing important structure in the data.

Bonus: Choose the Right Chart. Different graphs serve different purposes. Choose one that matches your task. Useful resources such as [Visual Vocabulary](#) guide on choosing the right chart⁷, while websites like [viz.wtf](#) showcase poor practices that serve as “negative inspiration”.

Main Takeaway from Part 1: Visualization is important. Moreover, the same set of underlying techniques is used to produce any visualization, independent of the fields. Whether you are showing population data, climate trends, or regression results, the visualizations can be produced and analyzed under a common framework.

4.2 Part 2. Visualization for Economists

The second part of this lecture turns the focus from general visualization principles to the specific needs of economists. In economics, we need to visualize not only our data but also our results. Graphical representations offer a clearer and more intuitive way to communicate both data and findings, thereby making research more transparent, engaging, and ultimately more impactful.

Contents of Part 2:

- [Data Visualization for Economists: The Problem & The Solution](#)
- [Visualize Data Before Estimating](#)
- [Visualize Econometric Results](#)

4.2.1 Data Visualization for Economists

1) The Problem and the Solution

Economists usually have an over-reliance on long descriptive statistics and regression tables, which are effective for detailed reporting. Yet, the problem is that such tables often overwhelm readers and obscure key insights. Moreover, we often write and present for other economists, neglecting broader audiences, such

⁷Section [4.2.2](#) elaborates on common charts used to explore data in economics.

as policymakers or colleagues from other disciplines who may not read tables the way we do.

The solution is simple: **we need to use visuals**. Visuals make our research more engaging. They make seminars more enjoyable, papers more readable, and referees easier to convince. An effective figure can replace several paragraphs of text and is often the part of the presentation that people actually remember.

Graphs can distill complex data into clear, accessible insights, aiding in the communication of findings to both technical and non-technical audiences by overcoming (some) communication challenges, as the general public does understand graphs better than regression tables. One additional point: Juniors sometimes tend to put all the results without a selective focus; this should be avoided. But at the same time, remember to document everything at least in an appendix.

2) Data Visualization vs. Visualization of Results

Think about visualization as playing two different roles in the research process: data visualization and visualization of results. Both stages are equally important. Visualizing the data guides the analysis; visualizing the results communicates it. In both cases, the idea is the same: to make information visible, intuitive, and credible.

- **Data visualization** comes first as an essential tool in economics before running any econometric analysis. It helps us understand the structure of the data, identify patterns and trends, spot outliers, detect errors, and even provides insights into relationships between variables. In the case of using complex data sets, visualization also facilitates communication of the data.
- **Visualization of results** comes after estimation for graphically presenting our main outcomes of econometric models. Ways to translate statistical estimates into clear, interpretable visuals include regression discontinuity and bunching graphs, event study plots for difference-in-differences, marginal effects graphs, etc.

Imagine an example like the aforementioned Anscombe's Quartet (Figure 8). Here, a regression table would clearly hide important information about the relationship between the x and y . In practice, there is no one-size-fits-all answer to

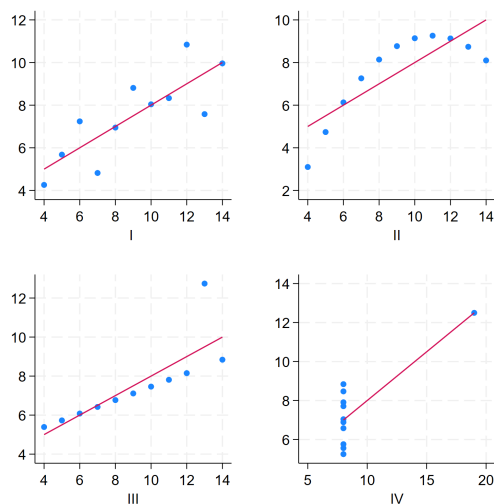


Figure 8: Anscombe's Quartet

the question of whether a figure or a table is better; it will depend on the specific context and purpose of the visual display. However, some general guidelines can help you decide when to use a figure instead of a table:

1. When showing trends or patterns over time or across variables, figures are often better for visualizing trends or patterns over time or across variables, using scatterplots, line graphs, and bar charts. Tables can also show trends, but figures can make them more visually apparent and easier to interpret.
2. Tables are often preferred to show alternative specifications or robustness checks.

Visualization is not a cosmetic addition to a project; it is part of the research process itself, as it helps us think and helps us communicate (and both are fundamental to doing good empirical research).

4.2.2 Visualize Data Before Estimating

Let's move to something even more fundamental: how we *see* the data itself. Before running any regression, before estimating any model, we need to understand what our data looks like.

1) How Do We See the Data?

Our brain processes visual information in two stages: a fast, automatic one and a slower, deliberate one. The first is called **pre-attentive processing**. It is an unconscious, automatic processing of visual stimuli that occurs before we consciously attend to them. We instantly recognize basic features of visual perception, such as color, size, length, and orientation.

The second stage, **attentive processing**, is the conscious processing of visual information that occurs when we focus our attention on a particular object or stimulus in our environment. It is slower and involves more complex cognitive processes such as object recognition, memory, and decision-making.

Visualization works because it takes advantage of this first, pre-attentive system. When data are encoded visually, we can detect patterns long before we have processed the numbers. The challenge is to design graphs that exploit these abilities without overloading them.

2) What Do We Want to Visualize?

The first step of data visualization is always to ask: *What type of information do I want to show?* Most answers fall into one of the following four categories:

- **Amounts:** how much of something exists.
- **Proportions:** how a total divides into parts.
- **Distributions:** how values are spread across observations.
- **Relationships or correlations:** how two or more variables move together.

After figuring out the type of information to show, we will proceed with choosing visualization tools. Below are some examples of using suitable tools for each category, drawn from the speaker's own research.

1. Visualizing Amounts

To show amounts, you can begin with a simple bar chart. For example, in a survey experiment on preferences for redistribution, respondents are asked whether differences between rich and poor are “not a problem”, “a small problem”, “a serious problem”, or “a very serious problem”. A bar

chart immediately shows which answers are most common. Adding a second dimension—for instance, comparing treatment and control groups— can reveal meaningful contrasts clearly (see Figure 9).

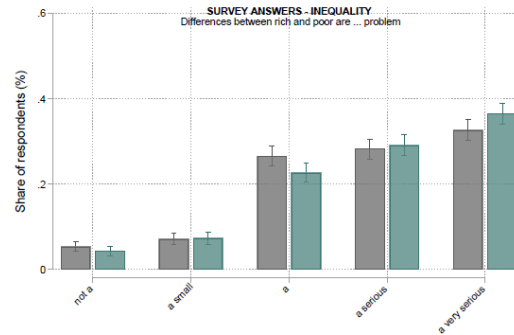


Figure 9: Example of Amounts Illustration (Foremny, 2024)

2. Visualizing Proportions

When interested in proportions, we often need to represent several categories within one total. Stacked bars are a simple option. Figure 10 (Foremny, 2024) shows how to plot regional variation in survey responses using horizontal stacked bars—one for each Autonomous Community in Spain. This makes it easier for us to compare how the perception of inequality differs across Spain.

However, with many categories within a variable, it can be difficult to illustrate the relationship between categories across variables. A further example comes from questions about identity: whether individuals feel they belong more to their nation or to their region (Foremny, 2024). Here, a **Sankey chart** is used to visualize how these two dimensions interact. Each flow connects regional to national identity, making it visually obvious that attachment to one often predicts attachment to the other (see Figure 11). This kind of plot helps us grasp multidimensional relationships that would be impossible to summarize in text.

3. Visualizing Distributions

Histograms have been a popular visualization option for showing distribu-

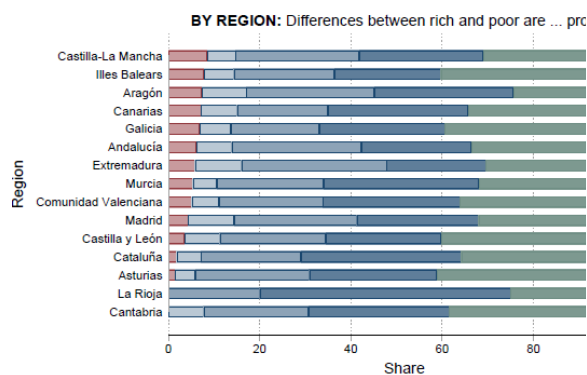


Figure 10: Example 1 of Proportions Illustration

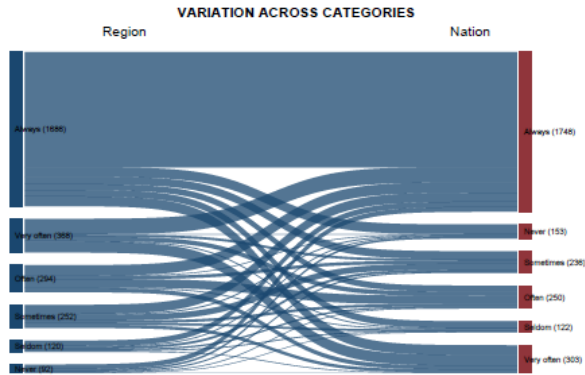


Figure 11: Example 2 of Proportions Illustration

tions since the 18th century. They depend on the bin width: a bin width that's too small may obscure trends in the data; a bin width that's too large may cause smaller features in the data to disappear. It's important to try different bin widths to accurately reflect the underlying data, rather than relying on the default option of any package.

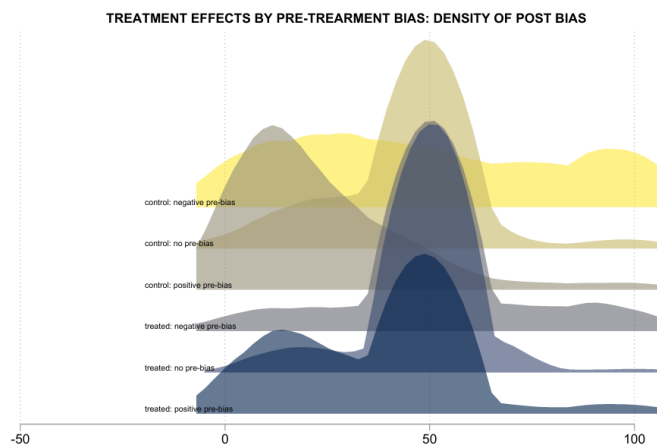


Figure 12: Density Plots

Nowadays, density plots are becoming more popular due to extensive computing power. They visualize the underlying probability distribution of the data with a continuous curve. Both allow us to see how a variable spreads across its range—where the mass of the data is and whether there are out-

liers. The curve is estimated using kernel density estimation, with a small width and a Gaussian kernel being the most widely used. Other kernel choices are available for density plots.

In fact, with visualization, **multiple distributions** can be analyzed by identifying a response variable and one or more grouping variables. Here, the response variable is the variable of interest, while the grouping variables divide the data into subsets with distinct distributions of the response variable.

The example above (Figure 12) comes from an experiment where participants were asked to estimate what share of taxes is raised by the central versus the regional government, with the treatment group provided information about the tax autonomy of regional governments (the true answer being 50%) (Foremny, 2024). Plotting the distribution of the treatment effects for different groups helps show if the treatment had an effect conditional on the pre-treatment bias.

4. Visualizing Correlations or Relationships

To visualize the relationship between two variables, it is very useful to use scatterplots or, increasingly, **binned scatterplots**. The `binscatter` is a simple yet powerful tool: Unlike traditional scatter plots, `binscatter` plots group observations into bins based on the values of one of the variables. The bins are usually equally spaced intervals, and the observations within each bin are summarized using a chosen statistic, such as the mean or median.

Take an example from Cattaneo et al. (2022). Figure 13a presents the usual scatter plot. Then in Figure 13b, the scatterplot is binned, and, within each bin, an average is calculated. Finally, Figure 13c presents the binned scatter plot, along with a polynomial fit.

Binscatter plots are particularly useful when the relationship between variables is nonlinear or when many observations overlap in a traditional scatter plot, making it difficult to discern patterns. By grouping the observations into bins, binscatter plots can reveal patterns that might be obscured in a

traditional scatter plot, especially when there is heteroscedasticity or clustering of the observations.

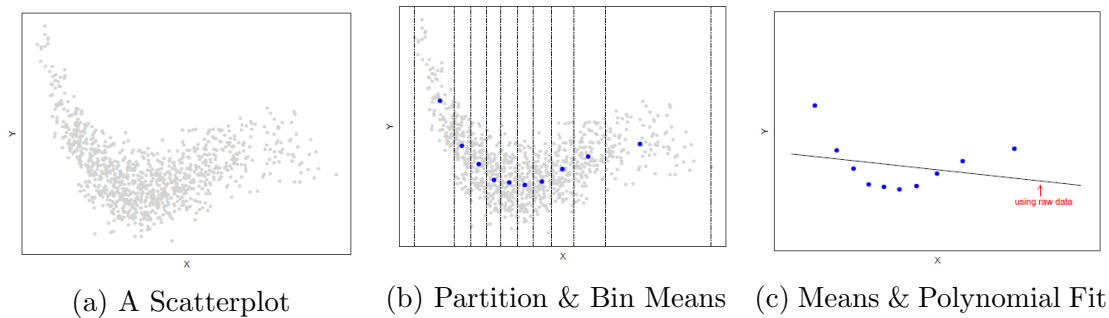


Figure 13: Example of Correlation Illustration

3) Some Basic Principles for Data Visualization

- **Keep it simple:** Avoid too many colors, shapes, or design elements that distract from the data.
- **Label axes clearly and accurately,** and Label each bar with the actual data value.
- **Order bars appropriately in logical order:** This might be by size, alphabetical order, or chronological order.
- **Use appropriate scales.** If the data varies widely, consider using a logarithmic scale.
- **Minimize clutter:** Remove unnecessary gridlines or borders, and keep enough space in between bars.
- **Use appropriate color:** Use easy-to-differentiate colors to highlight key information and differences.
- **Use annotations:** Titles, subtitles, and footnotes, or a data source.

4.2.3 Visualize Econometric Results

After discussing how to visualize the data itself, let's now turn to something that economists often overlook: how to visualize the *results* of our econometric models.

1) Tables and Graphs: A Balance

Historically, regression results were shown in tables. Traditionally, we have relied on tables to report dozens of coefficients, standard errors, and significance stars. Plots can make our results much more transparent, interpretable, and convincing, especially with large data sets, modern econometric tools, and larger computational power that together provide a good base for graphical visualization of econometric results.

Graphs complement, not replace, tables, in presenting results. There is a trade-off between using graphs and tables: graphs show more details, but tables allow for comparing different models and summarizing in a few parameters. Our goal is to **find a balance between the two**. In many cases, you can always show visuals in the main text, and report details of your estimates in tables in the appendix.

In the rest of this section, we'll review a few visualization tools that may be the most useful in applied research: **binscatter** and panel data, marginal effect and **marginsplot**, dynamic difference in differences and **coefplot**, robustness checks and **speccurves**.

1. Binscatter and Panel Data

As discussed in Section 4.2.2, **binned scatterplot**, or *binscatter*, provides a flexible, parsimonious way of visualizing simple relationships in data based on regression techniques. For presenting results, binscatter plots are also an important tool to visualize panel-data models, as it allows us to residualize on fixed effects and controls.

For example, Ferraz, Foremny, and Santini (2024) study the relationship between tax revenues and transfers in Brazil. Figure 14 shows how the binned scatterplots visualize different regression results, and help the understanding of how the relationship of interest changes as the regression evolves. The

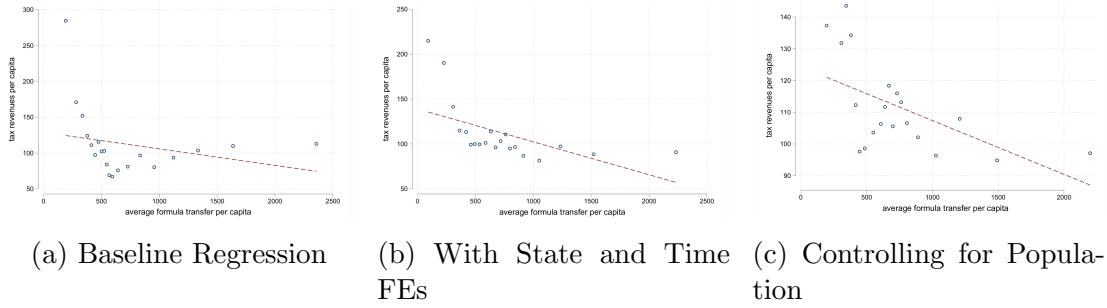


Figure 14: Using Binned Scatterplot for Panel Data Regressions

first subplot 14a presents the baseline regression, where there seems to be a negative and non-linear relationship. With state and time fixed effects, the non-linearity between X and Y reduces, as in subplot 14b. Finally, controlling for population resolves the non-linearity, as is shown in subplot 14c.

Here are some recommended practices when using binned scatterplots:

- Choose an appropriate number of bins: too few hides patterns, too many creates noise.
- Overlay the raw data or jittered points for transparency.
- Add fitted trends when appropriate.
- Adjust for covariates: consider residualizing variables if controlling for other factors.
- Label axes clearly and include units.
- Keep styles and scales consistent across related graphs.

2. Plotting Marginal Effects

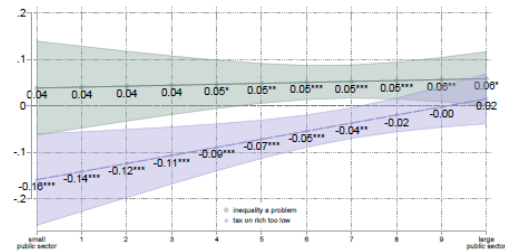
Plotting marginal effects is especially useful in nonlinear models (logit, probit) or whenever the model includes interaction terms and the effect of a variable depends on other covariates. In such cases, raw coefficients can be misleading because they represent partial derivatives evaluated at a point, and single reported coefficients alone do not capture how the effect changes across levels of interacting variables.

By plotting the predicted effect of a treatment or variable over the range of another covariate, we can see where the effect is strong, weak, or statistically insignificant. Confidence intervals provide a visual representation of uncertainty that is far more intuitive than asterisks in a table.

For instance, in Foremny (2024), the treatment dummy interacts with the perceived size of the public sector. Plotting the marginal effects of the treatment across the distribution of public-sector size reveals heterogeneity and illustrates where in the public sector size variable distribution the effects originate from (Figure 15).

Variable	inequality a problem			tax on rich too low		
	Coef	Std. Err.	P> t	Coef	Std. Err.	P> t
treatment	0.0382	0.0517	0.450	-0.1590	0.0503	0.002
size of public sector	0.0465	0.0051	0.000	0.0384	0.0048	0.000
treatment * size	0.0020	0.0072	0.780	0.0174	0.0069	0.012
Observations	2,812			2,812		
R-squared	0.0726			0.0777		
Adj. R-squared	0.0575			0.0627		

Results Table



Plotting Marginal Effects

Figure 15: Marginal Effects

Here are some recommended practices for marginal effects plots:

- Always visualize interactions or nonlinear effects.
- Include confidence intervals and coefficients in the plot.
- Plot over meaningful data ranges (avoid extrapolation).
- Highlight non-linearities or turning points.
- Label axes clearly and maintain consistent aesthetics.
- Combine plots with tables when necessary.

3. Event Studies

Event study graphs have become the standard way of presenting dynamic difference-in-differences results. They provide more comprehensive and nuanced insights into the impact of an event, allow for the examination of

the timing and duration of the effect, as well as any potential delayed or cumulative effects, and provide a test for pre-trends.

Here are some recommended practices for event study graphs:

- Normalize the event time: set the treatment period to 0 (except for the staggered approach).
- Show cumulative effects.
- Include confidence bands for transparency.
- For full transparency, include a graph (at least in the appendix) distinguishing treated vs control units, or pre- vs post-event periods.
- Use interpretable scales (percentage changes, deviations from baseline).
- Keep axes and styles consistent across related plots.

4. Specification Curve Analysis & Research Reproducibility

Finally, there is a powerful but still underused tool: the **Specification Curve Analysis (SCA)**, proposed by Simonsohn, Simmons, and Nelson (2015). It helps address one of the main challenges in empirical research: many published studies cannot be replicated, often due to insufficient methodological details. Empirical research allows many choices in data cleaning, variable transformations, and model specification, which can influence reported effects. Small researcher decisions (e.g., excluding observations, choosing covariates) can cumulatively favor a desired narrative. Instead of presenting just one preferred specification, SCA visualizes all of them.

Figure 16 shows an example from Ferraz, Foremny, and Santini (2025). In the specification curve, each model is represented by a point on the x-axis, ordered by some criterion (for example, the estimated coefficient of interest), and the y-axis shows the value of that estimate. Such a curve represents the distribution of estimated effects across a range of model specifications and provides **an overall picture of the sensitivity of the results to modeling choices**. Besides, if it is about including different controls and

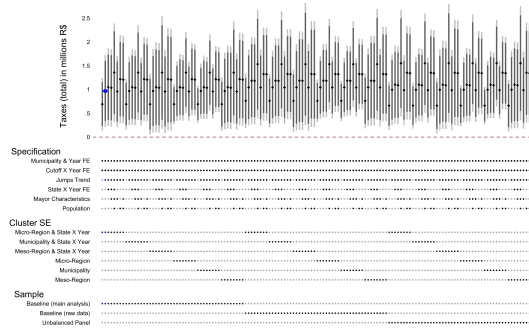


Figure 16: A Specification Curve

their impact, ordering by coefficient size can show **which variables or sets of them shift the point estimate** of the variable of interest.

The specification curve is very cool but rarely used in applied economics papers. We should use it more, as it helps with:

- (a) Robustness: By considering a range of plausible specifications, a specification curve can help identify which results are most robust and not highly dependent on a particular modeling choice.
- (b) Transparency: The curve provides a clear and concise visual representation of the sensitivity of the results to different model specifications, making it easier to communicate the findings to others.
- (c) Efficiency: The curve can help researchers to focus on the most robust and reliable results, thereby reducing the need for time-consuming and resource-intensive sensitivity analyses.

Here are some recommended practices when using specification curves:

- Define a clear set of plausible specifications in advance.
- Visualize all results, including those not shown in the paper.
- Report deviations from the planned analysis.
- Highlight influential choices and show which variables drive large changes.
- Use the curve to discuss robustness, not just to display it.

Main Takeaways from Part 2: Visualizing econometric results is not an optional embellishment; it is part of rigorous empirical practice. Figures help readers clearly see the data, the logic of identification, the patterns behind coefficients, and the robustness of findings. They make our work more transparent and reproducible, and they allow audiences outside academia to understand what our research means.

5 Lecture 4: Academic Talks in Economics

*Lecturer: **Jenifer Ruiz-Valenzuela**. Jenifer is an Associate Professor at the University of Barcelona and co-Director of the Public Policies Research Line at the Barcelona Institute of Economics (IEB). She is also a Research Associate at the CEP at LSE. Her research spans the Economics of Education, Labour Economics, and Public Economics. Known for her engaging communication style, Jenifer is a highly effective speaker in both academic and other public spheres.*

Presentations are a central component of the PhD journey, shaping how research projects are conveyed, evaluated, refined, and disseminated through seminars, workshops, conferences, the job market, and beyond. This lecture covers the various phases of preparing and delivering an academic talk, as well as how to benefit from one. It also extends beyond standard presentations to address other common formats, such as posters, discussant roles, elevator pitches, and interactions with the media.

Contents of this lecture:

- [Start Presenting Now](#)
- [Before the Talk: Design and Rehearse](#)
- [Deliver the Presentation](#)
- [Benefit from Academic Talks](#)
- [Other Presentation Formats](#)

Public speaking is a core skill for an academic. You need to **sell your research**, much like a vendor selling tomatoes at a supermarket. Presenting your findings to fellow economists or other interested audiences is an essential part of that process. While some academics are uneasy with this framing, selling in this context is not about exaggeration or marketing tricks. Rather, it is about giving the audience the information they need to understand the contribution of the paper and to see why the results matter.

Therefore, pursuing a PhD is not just about producing rigorous analysis and novel results; one must also obtain the ability to communicate them clearly, convincingly, and confidently to a wide range of audiences. The success of a project depends not only on its technical merits but also on how well it is conveyed. This lecture provides practical guidance on preparing, delivering, and benefiting from academic presentations throughout the PhD and beyond.

5.1 Start Presenting Now

It is never too early to start giving presentations. **Internal presentations**, such as “work-in-progress” seminars in your department, can and should start as soon as a researcher has a research idea. These early talks do not require polished results, let alone a complete paper. Their purpose is to practice explaining the idea, clarify the intuition behind the strategy, and receive feedback that can help push the project forward. Presenting unfinished work internally is one of the most effective ways to learn how others perceive the project and where confusion arises.

External talks, including conferences, workshops, and invited seminars (later in the career, of course), impose higher standards. In these settings, you need a story to sell, and the project should be quite polished, even if the analysis is not completely final. First impressions matter, and there is only one chance to give a first impression. For this reason, it is advisable to present externally only once the core idea, data, and empirical strategy are well developed and defensible. Over time, as you gain experience and confidence, these standards naturally become easier to meet.

Applying to conferences can be a useful commitment device, forcing progress toward a deadline. It is also a tool to examine the readiness of the work for external presentations. However, submitting work that is not sufficiently developed could be counterproductive. A good rule of thumb is: results do not need to be written up perfectly, but they should exist clearly and robustly in the analysis.

It is worth noting that Job Market Paper⁸ presentations require particular care. The appropriate level of polish depends on who is in the room. If potential employers are present, the paper must be highly refined, and you must be able

⁸More on this in Lecture 5.

to defend all aspects of the work. At the same time, Job Market Papers benefit enormously from being presented many times beforehand in “low-stake” settings, either internally or external student workshops, and with close coordination and feedback from advisors.

5.2 Before the Talk: Design Slides and Rehearse

When preparing for the talk, remember that you are a seller. People would want to buy your product if they feel they understand it and see its value. As [Rafe Meager](#) puts it: “You must give people the information they need to understand the results you will show them, and to see why the results matter.”

5.2.1 Slide Design: Principles

Slides play a supporting role in academic presentations. Their purpose is not to store all relevant information, but to guide the audience through the argument and to reinforce the speaker’s explanation. Well-designed slides reduce cognitive load, help listeners remain oriented, and make it easier to follow the logical structure of the talk.

A fundamental principle of slide design is **simplicity**. That is, each slide should convey a single idea as clearly as possible. Try to write at most one line per point. You should absolutely not clutter slides. Put as little information as possible to convey the point. Overly dense slides or long blocks of text quickly overwhelm the audience and make it difficult to understand where the presentation is going. As a rule of thumb, **if a slide feels crowded, it should be split into two**. If an idea requires substantial explanation, it should be delivered verbally rather than written out in full. Also, avoid typos!

When designing slides, visual emphasis should be used sparingly and consistently. A clear visual hierarchy, selective use of **bold** text, and restrained [use of color](#) help emphasize key messages without distracting from the content. Remember, emphasizing everything is emphasizing nothing.

Whenever possible, use graphs over tables for communicating results. Visualizations allow audiences to grasp patterns, magnitudes, and dynamics at a glance,

something that is difficult to achieve with dense regression tables⁹.

If tables are necessary, they should be adapted for presentation to be readable during the talk (from a distance without effort!). What you should *not* do is simply copy-paste the whole table from your working paper. Presentation tables should show only crucial information, such as the main point estimates, control summary ticks, and the number of observations.

5.2.2 Common Supporting Slides

Whether to include **an outline slide** depends on the context. In short talks like conference presentations, an outline is often unnecessary and can consume valuable time. In longer seminars or talks with unconventional structures, especially when having multiple methodologies or distinct sets of outcomes, an outline can help the audience stay oriented or highlight mechanisms. Besides, a brief preview of the main results can be useful in long talks, especially if there is a risk of running out of time.

For **literature slides**, the goal is *not* to provide a comprehensive survey, but to situate the paper in the current literature. Therefore, you should only include those very related papers. Highlighting how you differ from existing studies (i.e., your contribution to the literature) is far more important than listing many citations.

Since credibility is a primary concern for academic audiences, robustness checks are a common component of any presentation. Core robustness exercises should therefore appear in the main slides, integrated naturally into the narrative. At the same time, not all robustness checks need to be shown upfront. Additional analyses can be placed in backup slides and used selectively when questions arise.

When presenting models in empirical work, simplicity and clarity are crucial: assumptions should be stated explicitly, technical jargon should be minimized, and theoretical predictions should be clearly linked to the empirical analysis.

⁹More on this in Lecture 3.

5.2.3 Rehearse your Talk!

1) Why should I rehearse? Rehearsing a presentation is essential because it substantially improves the quality of the talk. From the perspective of the audience, preparedness shows that you take the presentation seriously and signals **respect** for the attendees. People do not want to waste their time sitting there for a poor presentation.

For you, the presenter, if you rehearse well, you will feel more confident and **less anxious** during the talk. You will rely less on the slides; you will read less and explain more, and engage more naturally with the audience. With enough rehearsal, you will also be able to use more precise language, making arguments clearer and more convincing. It helps identify weaknesses in the narrative, unclear transitions, or unclear intuition. Still, you will make mistakes, but fewer and smaller ones with rehearsals.

2) How can I rehearse? Rehearsing **by yourself** is the first and most important step. Timing the presentation, recording it, and observing one's own delivery can be extremely informative. You will be able to notice any distracting body language, like arms crossed and hands in pockets. It also helps to detect excessive use of filler words and unclear explanations. For non-native English speakers, rehearsal is particularly valuable for practicing the pronunciation of frequently used terms not natural in their native tongue. Video conferencing tools can be helpful at this stage, as transcripts can reveal repeated phrases or sections that are difficult to follow.

Practicing **with your PhD colleagues** is usually a helpful next step once you have mastered the individual rehearsal. Feedback from peers frequently highlights points of confusion that you may not anticipate and helps refine both content and delivery. Besides, as PhD students, providing detailed feedback for one another during these internal practices is a powerful way to foster mutual support and improve your presentation deliveries.

Presenting in **internal work-in-progress seminars** provides additional benefits. You would want to invite more senior researchers to attend your internal talk and hear what they think about your project. Constructive criticism from faculty members is an important input into improving the quality of the presentation.

Finally, the intensity of rehearsal should depend on the importance of the talk. For high-stakes presentations, such as Job Market talks, a stricter rehearsal discipline can be useful. Restarting the presentation from the beginning after a significant mistake reinforces fluency, sharpens transitions, and helps ensure a smooth and confident delivery.

5.2.4 Final Details before the Talk

Knowing the audience is critical. This is especially useful when designing the motivation slides. More specialized audiences, such as in a field-specific workshop, require a targeted motivation that dives quickly into the contribution, while a more general audience requires a broader “big picture” hook. Later in the career, when invited to give departmental seminars, it is often helpful to check the websites of the department and faculty members to understand the background of likely attendees.

Practical considerations also matter. First, clarifying logistical details in advance, such as the total time allocated, whether questions are taken during or after the talk, and how time warnings are given, reduces uncertainty and anxiety. It is also important to pay attention to dress codes. In general, you should neither be dressed too casually nor too formally. Try to strike a balance: avoid wearing either sweatpants or full suits (except for Job Market talks, where suits are less out of place). Lastly, in some talks, you might need to bring water, a pointer, and a backup of your presentation on a USB drive. These small details can all contribute to a smoother experience.

5.3 Deliver the Presentation

Calm down, you’ve got this! It is important to remind yourself that you are the person who knows the most about your paper. Before your presentation starts, depending on the setting, you can try to have some minutes for yourself, go to the toilet, and motivate yourself. In the first few minutes at the beginning of the presentation, you will feel the greatest nervousness. Here are some tips to handle this:

- Use the title slide wisely to calm down.

- Thank the person keeping track of the seminar and the audience.
- Briefly introduce yourself and co-authors.
- Explain in a couple of sentences what the paper is about.

5.3.1 Delivery: It's Your Stage

Do not start with an apology: try to inspire confidence. **Confidence** is conveyed through posture, movement, and tone. You should avoid standing still in front of the slides. Instead, move your body naturally, use your hands, and speak clearly and at a measured pace. A little reminder: do not use a pen that you can click on and off.

Academic presentation is a dialogue, not a monologue. You will need to **interact with the audience**. The basics are to look at your audience and talk *to* them, not at your slides or the air. In order to keep them engaged, make sure that you speak loudly enough so that people in the back of the room can hear you clearly without effort, and do not talk too fast. Also, be careful with jokes. A spontaneous light joke is fine, but you don't want to offend someone or make the talk unprofessional with edgy jokes.

Finally, be a careful tour guide for your project. Important concepts should be explained clearly from the moment you mention them; never assume a concept is self-explanatory. When figures or tables appear, you should actively guide the audience through what they are seeing and why it matters.

5.3.2 Handling Questions

Questions are an integral part of academic presentations. The mindset should be: Questions are good, they make your research better. When handling questions, make sure you first listen and then reply. Besides, it is often helpful to take brief notes on important comments (or ask someone to do it for you).

1) Listening. Questions should be approached with openness rather than fear. All questions are important; don't dismiss any. When listening to the question, you should look at the person and nod with your head, just to show that you

are actively paying attention. Questions should not be interrupted; wait until the question is over to reply.

2) Replying. Don't start with "It's a good point" in every answer. Even though you don't want to dismiss any question, not every question reflects a valid concern. If a question will be addressed in the upcoming slides, it is appropriate to say so and postpone it. When unsure about an answer, it is better to acknowledge it by saying that you need to think more carefully about it. Be careful with extra-long answers when you feel too comfortable.

Defending your work (especially your identification and model) is both legitimate and expected; argue *nicely* why you don't agree with certain points made by the audience, and do not get defensive. If there is criticism, it's of the project, not about you. It is possible to have a pushy attendee in the audience, you can always say "This is important, but if you don't mind, we can continue the conversation after the presentation," when they won't let a point go.

5.3.3 Managing Time Constraints

Use a watch or a timer on the phone to keep track of the time. You can also kindly ask the chair/organizer or a colleague to give you a signal when you have, say, 10 or 5 minutes left.

Even though you have prepared a lot, you might still run out of time. This can happen particularly if there are many questions. When this happens, the session chair or the seminar organizer might help pause questions for a few minutes so you can at least finish your main points.

If you realize you are not going to make it to the end, do not try to rush quickly through all your slides. Instead, focus on presenting the most important and convincing results, ensuring that the narrative of the paper and its central contribution are clearly communicated.

5.3.4 After the Talk

You're not done (yet) when the presentation is finished! To make the most of your presentation, **follow-ups** are necessary. As a natural extension of the talk,

you can try to reach out to the person who had any unfinished conversations or questions. Reach out also if you did not fully understand a comment. These can lead to productive exchanges or even future collaboration.

Writing down **a list of the comments** as soon as possible helps preserve valuable feedback that might otherwise be forgotten. In the notes, you would want to explain the content of the comment and write down your answers, or try to look for them if you did not have an answer. After presenting 2 or 3 times, many comments reappear, and by taking notes and reflecting on them, you will perfect the answer to those questions!

In some contexts, such as invited seminars or job market talks, it is appropriate and professional to send a brief thank-you email to the organizers.

5.4 Attend Presentations: How to Benefit

Presenting research is only one side of academic exchange. Attending seminars and workshops effectively is equally important for learning, generating ideas, and building professional relationships. Academic talks provide exposure to new questions, methods, and perspectives, and they are a central setting in which researchers learn how others think, argue, and communicate.

Simple habits, such as sitting in the front rows and minimizing distractions (put your phone away!), make it easier to focus during the presentation. Writing down some notes or key points helps understand the talk better (especially using old-fashioned notebooks). It is very natural to have doubts during the presentation. By asking questions, you can deepen understanding and contribute to the discussion, but it is important to pay attention and ensure not to ask repetitive questions.

Most departments and research groups have a weekly seminar in your (loose) field featuring invited speakers from an external institution. Engaging with these speakers outside the formal presentation is often just as valuable as the talk itself. As a PhD student, you should try to sign up and talk to the speakers. When meeting with a speaker, introduce yourself as a PhD in a specific field. You can often discuss a research idea (motivation, identification, results). Productive conversations may also focus on questions you have about a paper by the speaker that

you are interested in. Conferences and workshops also offer a great opportunity during breaks or meals to approach a presenter whose work interests you.

Over time, regular exposure to cutting-edge academic presentations also sharpens your intuition about what makes a research question compelling and how strong arguments are constructed, and contributes to the development of your own research agenda. In this sense, benefiting from academic talks is not only about absorbing content and talking with presenters, but also about learning how good research is done and communicated.

5.5 Posters, Discussions, and Pitches

In addition to standard talks, there is a range of other formats of presentation. While the objectives and time constraints differ, all formats benefit from clarity, focus, and an awareness of the audience.

1) Posters. It's all about the **visuals**. Poster presentation is a format in which you are more of a seller than in talks. Think of yourself again as a vendor standing in the market with your product while your potential audience passes by. Since interaction is often brief, you should choose wisely the main result and the main mechanism to include in your poster.

At the beginning of the poster session, probably no one is stopping by. WHAT TO DO? You should try to make eye contact and smile at people walking by. Very likely, someone will then stop at your poster. To break the ice, you can start by asking whether they are interested in learning about the impact of X on Y. Again, very likely, people tend to be nice and say yes! If you manage to do so, the “contagion” starts, and other people will also stop by.

2) Discussions. Serving as a discussant requires a positive attitude and a constructive mindset. You don't want to be the toxic element in the room! To make a good discussion, it is advisable to prepare a few slides, if allowed, or some clean and organized notes. This typically starts with a very brief summary of the paper before moving on to the discussion part, which typically includes:

- **Pointing out the nice aspects.** These could be, just to name a few,

the contribution, the great data, innovative research design, or unexpected policy implications.

- **Your contribution.** You can flag potential robustness checks, suggest additional analyses (e.g., other outcomes or likely mechanisms), mention literature that could help the authors, discuss policy implications that were not mentioned, etc.

Throughout, it is important to respect the allotted time and maintain a collegial tone. Remember, this is not *your* show, but to help others improve their projects.

3) Elevator pitches. This format is an extreme test of clarity. Pitches typically last no more than 30 to 60 seconds, or 100 to 150 words, if you draft them. You should avoid technical language and jargon at all costs. An effective elevator pitch:

- grabs the attention immediately by starting with a clear and engaging research question,
- explains what you are doing and why the question matters,
- states the main result in simple terms (answer to the question),
- ends with a hook or next step –something memorable or actionable.

Practicing elevator pitches with non-academic audiences (your parents/grandparents, friends who are not economists, etc.) is a useful way to ensure that the core message of a project is truly clear.

5.6 Bonus: Talking to the Media

Engaging with the media is a whole other talk, even though some of the points in the previous sections are still valid. Media formats are often unfamiliar, time-limited, and less forgiving than academic settings. When contacted by the media, ask for questions in advance. And then you will prepare a lot for it. If you get any question unprepared for, bridge unexpected questions back to prepared messages (e.g., “This is a good point, but what I can tell you from my research is . . .”).

Different formats, different strategies. Bring notes for recorded interviews and radio. Television appearances benefit from practicing body language and avoiding distracting clothing. Regardless of format, the guiding principle remains the same: clarity, preparation, and control of the message. The last piece of advice is always to ask any questions beforehand: time, microphone type, standing/seating, etc.

This lecture discusses how to give an academic presentation and how to benefit from one. It is never too early to start presenting. When preparing for a talk, use a simple slide design and rehearse as much as possible. During the talk, be confident, interactive, and professional, and remember to handle questions with openness. Beyond presenting, junior researchers benefit from attending talks to shape good academic taste. Finally, mastering other formats of dissemination ensures research is effectively communicated throughout the PhD and beyond.

6 Lecture 5: Crafting the Job Market Paper

*Lecturer: **Albrecht Glitz**. Albrecht is an Associate Professor at UPF and BSE, a Project Leader at RFBerlin, and a Research Fellow at CEPR, IZA, CESifo, and CReAM. His research spans Labor economics, migration, and the economics of espionage. Having supervised numerous PhD students and served repeatedly on hiring committees, Albrecht brings first-hand insight into how to craft a well-positioned Job Market Paper that signals the ability to produce independent and high-quality research.*

This lecture focuses on how to *write* the most important research product during the PhD journey: the Job Market Paper (JMP). The JMP is by far the most important item considered by prospective employers in the job market.¹⁰ During the PhD, one should focus their research efforts on the JMP. As a PhD Candidate entering the market, it is more important to have one finished JMP than many papers at earlier stages. In principle, the JMP should look like a paper in a top economics journal and should be written at the highest professional level for a wide audience of highly trained economists. This lecture is also helpful for writing other papers besides the JMP.

Contents of this lecture:

- [The Job Market Paper](#)
- [The Structure & Writing Guidelines](#)
- [Writing and Rewriting](#)
- [Final Remarks](#)

¹⁰The primary focus of this lecture is writing. For generating a promising idea and selecting a project for the JMP, see Lecture 1 in Volume 1, as well as the Opening Lecture and the Closing Lecture.

6.1 The Job Market Paper

While the PhD dissertation typically includes three related or unrelated papers as its chapters, the JMP is more than just *a chapter* and must stand alone as *one self-contained, polished* paper, written as if it could be submitted to a top journal.

The primary purpose of JMP is to show that you can carry out a significant piece of independent research. It should be your most important written research product on the market; it should demonstrate originality, rigor, independence, and taste; it serves as a signal of your long-run research agenda. It is probably the only document that employers actually read before flyouts!

When departments receive hundreds of applications, committees review dozens of applications under tight time constraints, and the Job Market Paper becomes the decisive factor in determining which candidates advance to interviews. Furthermore, departments read the JMP not only to evaluate the specific project but also to infer your trajectory as a researcher: what you have been able to do during your PhD is interpreted as a signal of what you will do in the years ahead.

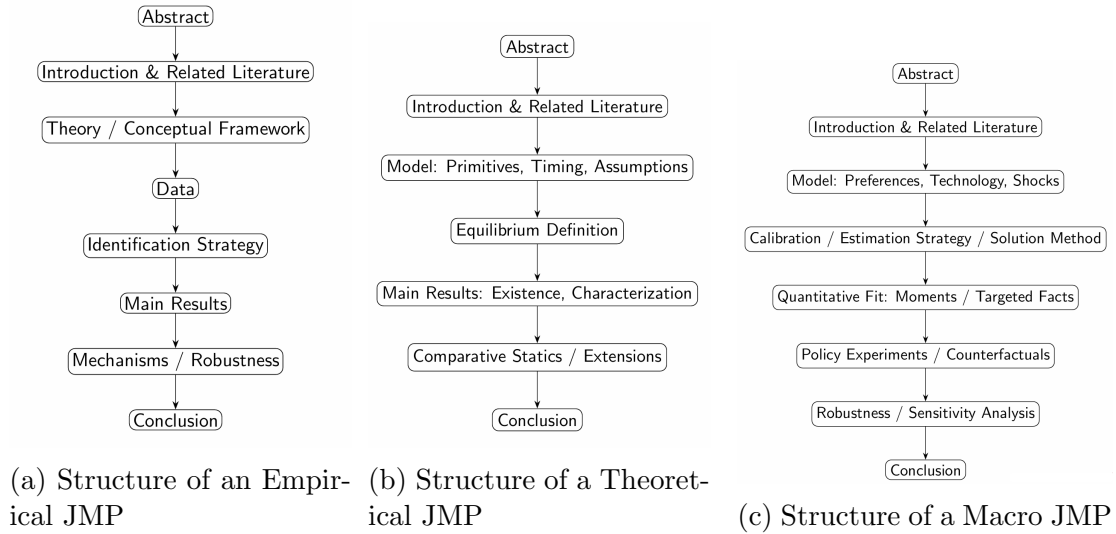
A successful JMP therefore has to have **five elements**: a clear and important question, a credible identification strategy or tight theoretical logic, a meaningful contribution relative to existing work, a clean exposition and professional presentation, and evidence that you can generate a stream of future papers.

6.2 The Structure & Writing Guidelines

Like all other research papers, a JMP includes a title, an abstract, an introduction, and a conclusion. The typical empirical and theoretical JMPs then differ in emphasis in their body parts. While this structure is typical, it's not set in stone. Researchers may add sections when necessary (such as case studies), but should have good reasons for deviating from the standard format, as any deviation will draw attention from readers.

An empirical JMP (Figure 17a) typically begins with an introduction that lays out the research question, motivation, and contribution. It then provides a conceptual framework, along with the necessary context or institutional background. Following the description of the data, it describes the identification strategy and presents the main results. Subsequent sections explore mechanisms, heterogene-

ity, and robustness, before concluding with the broader implications. Technical details, proofs, and data cleaning are usually placed in an appendix, keeping the main text clean and accessible.



A theoretical JMP (Figure 17b) follows a similar logic but places more weight on the modeling section. After motivating the question and explaining its relevance, the paper typically provides an intuitive overview of the environment and key mechanisms before introducing the formal model. The formal section defines primitives, timing, assumptions, and equilibrium definition and characterization; subsequent sections develop the main results and comparative statics. Macro and structural JMPs (Figure 17c) typically feature reduced-form empirical evidence, structural parameter estimation, model calibration, quantitative fit assessment, and counterfactual analysis. As with empirical papers, lengthy proofs belong in the appendix, while the main text focuses on clarity and intuition. Regardless of type, the unifying objective is clarity: A reader should never feel lost about the structure of the argument or the purpose of each section.

6.2.1 Crafting the Abstract and Introduction

The abstract and the introduction are, by far, the most important parts of the Job Market Paper. They are the entry points to your work, the basis on which hiring committees form their first impressions. A strong JMP, therefore, should

begin with a precise, disciplined, and persuasive opening.

1) The Abstract The abstract must communicate the essence of your paper in roughly one hundred words. Its role is not to summarize everything you do, but to make the reader understand immediately *what the paper is about, what you do, what you find, and why it matters*. A well-crafted abstract typically follows a clear structure:

- Start with the **research question**, stated plainly and without jargon.
- State **what you do**, mentioning the empirical strategy, theoretical framework, and setting.
- Present the **main results with magnitudes**: numbers matter here.
- Close with the **key contribution and implications**.

Keep the abstract concise. It should be **5 to 7 sentences, or around 100 words**. It should contain no vague statements. Background or literature lists should not be included, as the abstract is not the place to “set the stage” but the place to communicate content. Because of this, it should be written last, once the story of the paper is fully clear in your mind.

Exercise 1: Analyze the abstract of your JMP (or one by someone else) according to the discussion above.

2) The Introduction The introduction is **the most important section** of the Job Market Paper because it’s often the only section that recruiters will read *in detail* before a flyout. Besides, the introduction is often the only section of a paper that people actually read after publication. Therefore, it needs to be edited and rewritten many times, and **should be “perfect”**. [Keith Head](#) provides a good structure for the introduction:

1. **The Hook.** Open with a tension, a puzzle, or a stylized fact that makes the reader care. This is not decoration: it is your opportunity to demonstrate the economic relevance of the problem. (1 paragraph)

2. **Research Question.** State clearly what the paper does. A sentence such as: “*In this paper, I examine the effect of...*” is often the most effective. Avoid vague descriptions or overly broad claims. (1 paragraph)
3. **Why It Matters.** Explain the economic importance of the issue. This is where you situate the question in the broader policy or theoretical context. A reader should immediately understand the stakes. (1 paragraph)
4. **Literature Gap.** Identify what previous work has not addressed. A typical formulation is: “*While prior work has examined X and Y, we still know little about Z*”. This gap creates space for your contribution. (1 paragraph)
5. **Contribution.** Present your contributions clearly and explicitly. The standard format remains the most effective: “*This paper makes three contributions.*” Then list them in a logical order, each tied directly to the gap identified earlier. This paragraph tells the reader why your work should exist. (1 paragraph)
6. **Preview of methods and data.** Offer a concise roadmap of your empirical or theoretical approach. This is the place to explain *how* you answer your question: which data, which variation, which model, and which key assumptions. Aim for clarity rather than technical completeness. (1-3 paragraphs)
7. **Preview of the main results.** Present the central findings with magnitudes and economic interpretation. Numbers matter here: they anchor your contribution and give the reader a reference point for the rest of the paper. (1-2 paragraphs)
8. **Literature.** Nowadays, literature is rarely isolated in a separate section; instead, it is integrated into the introduction and revisited where needed in later sections.
9. **Optional roadmap.** Most papers end with a brief roadmap of the structure (“*The rest of the paper proceeds as follows*”). This is optional: many readers skip it, but it can be helpful if the paper is long or conceptually complex.

***Exercise 2:** Compare two versions of an introduction paragraph, and tell why the second is better based on the discussion above.*

Version 1: Unemployment benefits are an important policy tool. In this paper, I discuss various aspects of unemployment insurance and present some empirical results.

Version 2: This paper asks how unemployment benefits should vary across regions that differ in wages and housing costs. Using administrative data and a quasi-experimental reform in Germany, I show that a more spatially differentiated benefit system can improve welfare by X%.

It is worthwhile to mention here the **contribution and positioning** of the paper. The recommended practice is to integrate the literature throughout, not only in one section. Strong JMPs weave positioning throughout the introduction and return to it whenever relevant in later sections, helping the reader understand continuously how the argument connects to and improves on what we already know.

You must clearly state **what is new** in the paper and why that novelty matters for the broader field, and explain **why your approach improves on existing work**. The contribution should be tied directly to the identified literature gap.

Precision matters when explaining the contribution. Overselling weakens credibility — readers quickly detect exaggerated claims or sweeping statements that cannot be supported. Underselling, however, obscures the value of your work and leaves the reader unsure why the paper should exist. Effective positioning strikes a balance: confident but careful, ambitious but precise.

6.2.2 Writing Guidelines for Subsequent Parts

For each of the other parts of the Job Market Paper, there are some key guidelines, both for keeping in mind when writing the first draft and as a checklist when revising the paper before releasing it.

1) Models and Theory For theoretical or structural JMPs, clarity in the modeling section is crucial. Before presenting any formal derivations, the reader must have a good understanding of the setup. This requires stating clearly the

primitives, timing, and assumptions. It is essential to provide the economic **intuition** before formalizing. Notation should support the argument, not obscure it. A clear theoretical section demonstrates not only technical competence but also your ability to communicate complex ideas effectively.

A well-written theory section offers a clear **roadmap of the model**, not just equations. The reader should understand why each assumption is made, how the components of the economic environment fit together, and what mechanisms the model is designed to isolate. Technical proofs and long derivations belong in the appendix, allowing the main text to highlight the economic **mechanisms** that drive the results.

It is also worth noting that employers nowadays increasingly value seeing a (simple) theoretical framework even in primarily empirical papers, as it demonstrates that the author can think theoretically about their empirical analysis.

2) Data Section The data section should demonstrate that you take empirical work seriously and are meticulous in your approach. It should give a clear and coherent picture of the empirical environment, explaining in detail where the data come from, and what period and geographic unit they cover. Explain **sample construction** and selection criteria carefully and precisely: what observations are included or excluded, how **key variables** are defined, and which transformations (if any) are applied.

When presenting **descriptive statistics**, it is often helpful to show the distribution of the main variables — through percentiles, density plots, or simple visualizations — so that the reader can understand the range, variation, and potential sources of heterogeneity in the data. These elements give context to the identification strategy and help the reader interpret the magnitude of the results. Descriptive statistics are also useful for us as researchers to identify potential data issues, such as detecting outliers. You should show variables as they appear in your analysis (e.g., changes rather than levels if your regressions use changes).

Detailed coding decisions, cleaning steps, or complex construction procedures should be clearly documented, but reserved for an appendix or replication package to keep the main text accessible. By the end of this section, the reader should have complete confidence that the empirical analysis is built on a solid and transparent

foundation.

3) Identification The identification section is the backbone of any empirical Job Market Paper. This is where you show the reader not only what you estimate but why the estimates can be interpreted causally. A strong identification section first describes, with clarity and precision, the **source of variation** that identifies the parameters of interest, be it a reform, an instrumental variable, a discontinuity, or an unexpected shock.

A compelling identification section also requires stating your **identifying assumptions** explicitly. Then you should explain why assumptions are credible with supporting evidence. This is where balance checks, pre-trends, falsification, and placebo tests play a role. The goal is not to overwhelm the reader with tests, but to offer a set of targeted diagnostics that speak directly to the assumptions of the research design. If you use an IV, you should always report first-stage results to support its relevance.

The last critical step in writing this section is acknowledging **limitations**. Acknowledging them shows maturity as a researcher. Every empirical strategy has weaknesses, and readers expect you to show that you have thought about them and explain how you address potential concerns. Often, limitations that seem fatal to young researchers are accepted by experienced readers if addressed honestly.

4) Presenting Results Once the empirical strategy is established, you will write your results to build a **coherent narrative**, with each table or figure advancing the story. The first step is to introduce the **main specification** and the core coefficients. After the main results, the narrative should unfold naturally. Mechanisms, heterogeneity, and alternative explanations are explored in separate but logically connected subsections.

When writing the results, bear in mind that the emphasis should always remain on providing the **economic magnitude** of the estimates, not simply reporting numbers and statistical significance without translating them into meaningful economic terms. You should interpret signs, sizes, and mechanisms. Readers want to understand how large the effects are, why they arise, and whether they

differ across relevant groups or contexts. If coefficients come with unexpected signs, they should be addressed, not ignored. If there are possible alternative explanations, you should address them systematically.

A key suggestion in crafting the results section is to use clean, intuitive figures and tables. A good JMP should have figures and tables that can stand alone with comprehensive notes. They should be replicable from the information given, and fully labeled and clearly titled. **Visualization** greatly enhances this part of the paper, and a good figure can make a significant difference in how readers perceive your paper¹¹. When using figures, the principle is to use clean and **intuitive visual design** and minimize clutter.

Tables should be used in a smart way: they should highlight the key coefficients rather than display every robustness check or specification. Avoid tiny fonts or too many decimals. Titles of the tables should state findings, instead of being just “Table 3”.

5) Mechanisms and Robustness A credible JMP does not stop at estimating a main effect; it seeks to explain the underlying mechanisms. This requires designing empirical tests that speak directly to the channels implied by the conceptual framework or the theory. Mechanism exercises help readers understand not only *whether* an effect exists, but *why* it arises and under what conditions it might be expected to appear.

Robustness checks complement these exercises by showing that the results are not driven by artifacts. Strong papers approach robustness strategically: they test a focused set of plausible alternative specifications and address specific concerns with targeted tests, rather than using “kitchen sink” robustness checks.

A good strategy to keep in mind in writing this part is to try to put yourself in the position of a potential referee, anticipate their concerns, preempt the 3-5 most likely referee objections, and then address these proactively in your paper. This is also to demonstrate that, as a strong job market candidate, you have thought about potential criticisms.

¹¹See “Lecture 3: Visualizing Economic Research” for a more comprehensive review on this matter.

6) Conclusion Many conclusions simply repeat what was already explained in the paper. This should be avoided. A strong conclusion focuses on key takeaways, explains why they matter, and situates them within the larger debates or policy questions of the field. This is also the place to discuss the implications of the work — whether theoretical or policy-relevant — and to acknowledge limitations honestly. Besides, a brief but thoughtful suggestion for future research gives the paper a sense of openness and direction. The best conclusions leave the reader with a clear sense of contribution, relevance, and intellectual coherence, reinforcing why the project deserves to stand as a Job Market Paper.

6.3 Writing and Rewriting

In the end, writing a JMP is about communicating your findings effectively (in English). Good writing will help the reader identify the precise contributions of your paper, making life easy for your reader. Bad writing deters readers from finishing reading your JMP. But more to that, bad writing will be used as a negative signal about your skills as a researcher, since it often accompanies muddled thinking. This section first presents some language guidelines for writing the JMP, which are especially useful for non-native English speakers, and provides insights on how to improve the writing once you already have a draft.

6.3.1 Writing Style: Language Guidelines

A good writing style in economics rests on a few principles. First, an effective paragraph contains only **one idea** and follows a simple structure. It should begin with a **clear topic sentence** that clearly states the main point or claim that the paragraph will develop. A good strategy when writing a paragraph is to ask yourself: If I read only the first sentence, do I know what the paragraph is claiming? If not, the first sentence needs to be rewritten.

Second, **use short, direct sentences**, and avoid “throat-clearing” phrases. Broad and vague openings such as “*In recent years, there has been growing interest in...*” or “*The issue of X has become increasingly important...*” add length to the sentence without adding any information. Also, use active voice, as it makes sentences more direct and easier to understand (Use “I show...” rather than “It

is shown. . .”).

Finally, at the word level, you should always **remove unnecessary and filler words** (indeed, very, somewhat). If you can take a word out without changing the meaning, take it out. Pay attention also to possible repetitions and avoid them. Besides, choose simple words over complex ones when they convey the same meaning. You should **avoid acronyms and jargon**, especially in the introduction. A reader encountering your work for the first time should not have to decode terminology before understanding the question. But whenever necessary, define terms immediately when used and **stick to the terminology**.

Here is an example of a Jargon-heavy start: “Using a spatial RDD with a Bartik-shift instrument and an AKM value-added framework, I estimate the LATE of benefit variation on equilibrium outcomes.” A clearer version would be “Unemployment benefits are the same everywhere, yet living costs differ dramatically across regions. This paper asks whether benefits should vary with local prices.”

6.3.2 Rewriting as a Process

Good writing is rewriting, so expect many revisions. It is all about “molding a product” that changes constantly until it finally settles into its final form. Think of rewriting as a process. You should plan for multiple full passes through the paper: full re-evaluations of structure, flow, and clarity. Note that the introduction should be rewritten last, once the rest of the paper has stabilized. Another useful practice is to have your peers review your paper (and be tough with feedback).

During the rewriting process, you will cut many pieces from the previous draft. It is important to maintain a “cut file” for removed text that might be useful later. Version control systems (even manually creating monthly copies of entire project folders) or a disciplined use of tracked changes help prevent chaos as the paper evolves.

Presentations at internal workshops and seminars¹² are opportunities to identify unclear aspects of your JMP that need improvement. If seminar audiences consistently get confused and ask clarifying questions about a particular part,

¹²More on this in Lecture 4.

that section of the JMP needs revision first.

Common Mistakes Previous sections have set up good practices in writing the Job Market Paper. However, certain pitfalls recur frequently across JMPs. Reviewing them makes it easier to avoid them:

- **Inflated motivation without a clear economic question.** Long introductions with broad statements or policy background but no sharp research question leave readers unsure what the paper is actually about.
- **Literature “dumping.”** Excessive literature review that does not highlight your contribution. Again, strong JMPs “integrate” the literature to highlight the gap the paper fills.
- **Overly technical exposition.** Pages of algebra, formal notation, or econometric detail can overwhelm readers and obscure the underlying intuition. Technical material belongs in the appendix unless it is essential for understanding.
- **Inconsistent variable names or notation across sections.** Using different names for the same concept across sections confuses readers. Stick to your defined terminology.
- **Burying the main result.** When the key finding appears for the first time on page 25, the impact of the paper is lost. The main specification and central results should be presented early and clearly.
- **Poor writing and repetitive content.** This signals muddled thinking to employers. Spelling mistakes and grammatical errors suggest carelessness.

6.4 Final Remarks

The “Goldin–Katz ten rules” are a helpful summary of what makes a strong Job Market Paper, and echo with many of the points raised in this lecture:

1. It is always possible to transform a good idea into a great JMP.

2. The insights of your JMP will be judged by how you present them.
3. Your JMP is an exercise in persuasion.
4. No great paper first emerged from the author's printer in that form.
5. No author can see all (or even most) of his or her writing errors.
6. Most paragraphs have too many sentences, and most sentences have too many words.
7. The "foot-in-the-door" parts of your written work are the abstract and introduction.
8. Verbalizing your argument is more difficult than writing it.
9. Be your own worst enemy.
10. There are many other rules: See the Goldin and Katz (2009) original [pdf file](#).

Remember, your JMP is the most important component of your job market application. Treat it as your flagship project: choose a question you care about, develop it with rigor, and present it with clarity and precision. Clear structure and careful exposition are as crucial as the substance of your contribution. Rewrite repeatedly and seek honest, critical feedback. You should aim for a paper that a busy economist *wants* to finish reading.

7 Closing Lecture: Research Products in Economics

*Lecturer: **Diego Puga**. Diego is currently a Professor at CEMFI and a leading scholar in urban economics, economic geography, and international trade. His research has been published in top academic journals and has also served in editorial roles at several major journals. Beyond academia, Diego has been deeply involved in policy work at both national and international levels, and is currently a member of Spain's Council of Economic Advisers.*

Throughout this *Guide*, we have learned how to navigate the research process, generate and refine ideas, organize the workflow, visualize findings, deliver presentations, and craft a Job Market Paper. This closing lecture ties these threads together by shifting the lens to building research products. A paper is one product among several; elevator pitches, presentations, job market talks, and policy briefs are all distinct outputs that a researcher must learn to craft. What unites them is that every product must convey a story based on credible research. This lecture first addresses how to select a research project promising to turn into a research product. It then covers various academic products according to their length (short versus long) and medium (written versus oral), showing how the same core elements scale and adapt across formats, and briefly discuss policy-oriented products.

Contents of this lecture:

- [Before the Product](#)
- [Academic Products: an Overview](#)
- [The Short Academic Product: Abstract and Elevator Pitch](#)
- [The Long Academic Product: Introduction and Seminar Spiel](#)
- [Policy-Oriented Products](#)

7.1 Before the Product

Building a research product starts with having a research project, but not every project translates into a compelling product. Before discussing various research products in economics, it is worth asking what characteristics make a project more promising than others and what we should do as projects evolve.

7.1.1 Selecting a Project

Be creative and daring. As a PhD student, your most important research product will be the Job Market Paper. From the experience of evaluating many job-market candidates, hiring committees are far more impressed by candidates who have a novel JMP, where they have tried to do something different, even if it is maybe rough on the edges and less polished, than by someone with a paper that is technically polished but makes only a marginal advance.

Of course, difficult things can fail more likely. Trying to tackle a difficult question and succeeding only in part can still demonstrate originality, courage, and the potential to do important work, which can already be enough to earn a flyout. Even a completely failed project is a learning process. The person willing to tackle new and difficult things is someone who will sooner or later do something important. And that's exactly what is valuable to show for a young researcher.

To have a creative project, start with an **actual research question**¹³, not with a method or a dataset. It is tempting to begin from a method you want to apply or a dataset you have gained access to, and that is fine as a starting point, but it is usually much more productive to let the question come first and then search for the tools and data that best serve it.

Your motivation should come primarily **from the real world** rather than the literature. Typically, it would be things that interest you, and people really notice when you genuinely care about the research, the question, and the answer, or just doing something mechanical. Saying “people have done X and I will do it differently” is fine, but saying something concrete about the world and why it matters is always more compelling. Caring is two-sided. It is partly about

¹³The following suggestions complement those in the Opening Lecture and Lecture 1 on generating ideas.

excitement and partly about investing deeply: reading, thinking, and learning every detail of your context.

This requires that you develop **deep knowledge of your context**, to know the ecosystem, and invest in that. Especially in applied fields like development, economic history, and labor economics, there is a tradition of researchers being deeply invested in institutional details. Superficial approaches like “I have this data and these regressions, but I have never been to that country” are increasingly common and increasingly penalized.

7.1.2 Rethinking your Project: Mind the “History Trap”

Most research projects change substantially over time. The initial idea, the initial prototype of a project, is hardly ever the best. There is a solid reason behind: good research is about discovery. If you knew from the beginning exactly what you would end up doing, it probably would not be that surprising.

We all acknowledge that progress usually comes from adjusting the approach, redefining the focus, or rethinking key choices along the way. However, probably **the single biggest mistake** researchers make is letting your product reflect the history of how you did the project. At some point, it is useful to stop refining the existing draft and instead rethink the paper from the reader’s perspective, asking how the story should be told if it were written from scratch.

The antidote is a simple question: *This is my result, this is my story. What is the best way to communicate this if I were to do it from scratch?* And sometimes it is actually good to do it from the ground! For example, when writing a paper, this sometimes means setting the old introduction aside entirely, sketching a new structure, and only then recovering pieces from the previous version.

This is important because it is easy to get trapped by the history of the thing. It’s easy to think about it only from your own point of view as someone who researched, as opposed to someone who is coming in for the first time and trying to understand what you’re up to. Therefore, keep an open eye and keep rethinking the project. Always ask yourself from time to time: “Is this the best thing I can do with this? Is this the right way to do it?” Always think of people who have seen this before: what’s the best way to tell them the story?

7.1.3 General Advice for Junior Researchers

Before diving into different research products, here are some additional suggestions. First, at this stage of the career, it's important always to keep a balance between two things: **breadth and steady progress**. Intellectual breadth is important at this stage. Reading papers, attending seminars, and taking courses outside of your core area can generate ideas through unexpected connections across different things.

At the same time, be realistic. Writing a good paper takes time. Eventually, you will finish your PhD and need to look for a job. By then, you will need a product to sell. Most researchers discover the “right version” of the paper through trial, error, and iteration. Besides, if you have some margin, even just several extra months, there will be time to really communicate your product and have it in a shape that people can understand.

Finally, **seek opportunities to collaborate**. Working with other students or with senior faculty is enjoyable, educational, and exposes you to different ways of thinking. Early connections are important because they are likely to last for your career. You keep seeing the same people for years, and you work together, sometimes you don't, but you still keep in touch and collaborate.

7.2 Academic Products: an Overview

Every research product tells a story, and every story is told in multiple formats: the paper itself, seminar presentations, meetings with advisors, elevator conversations at conferences, poster sessions, meetings with visiting speakers, and eventually job-market interviews and talks. Each of these academic research products has its own special advantages and constraints, but they also share many elements. They differ along two dimensions:

- **Length.** A *short product* conveys the essence in under a minute of reading or speaking (roughly 100 words). A *long product* takes several minutes to read or deliver (roughly 2–3 pages, or the first few minutes of a talk).
- **Medium.** A *written product* (abstract, introduction) is polished prose that must work on the page without the author's voice. A *verbal conversational*

product (elevator pitch, seminar opening, job market spiel) must work in real time with interruptions, body language, and audience interaction.

These two dimensions produce four cells:

	Written	Oral
Short	Abstract	Elevator pitch
Long	Introduction	Presentation Opening / Job Market Talk

The short and long versions are the primary axis. Within each, the written and oral versions are structurally parallel: the abstract is essentially the *script* for the elevator pitch, and the introduction is the *blueprint* for the seminar opening. They receive disproportionate attention compared to other parts in the paper or the presentation, because they often determine whether an audience engages further. Many principles for the written product extend naturally to oral and informal communication. Mastering one helps you master the other. The following sections take each row in turn, before briefly discussing some principles of the full, longest version: the paper. Later, we will also explore a parallel category, policy-oriented products, which sits outside the academic grid.

Remember, these products are all about **communicating your research**. Some researchers treat communication as secondary, as “marketing” rather than “real research”. But communicating research is not a secondary or superficial task. It is an essential part of the research process. You are selling something you care about, something you trust, something you believe is good. If the author does not articulate that story, no one else is going to do it for you, and all of those months and years of effort working on a project will be lost.

7.2.1 Start Early, Revise Often

Take the written format as an example. When you have your tables and figures, or your model and propositions, getting some words around that is relatively straightforward. Yet writing the abstract, the introduction, and the conclusion is much more difficult to do. But they are key because those are what people read first, and if you don’t do it well, it will also be what they read last.

For this reason, it is important to devote time to drafting them right from the beginning. Even if they don't say what the results are because we don't know yet, the motivation is there, the question is there, and the methodology is there. This early draft serves a double purpose: it forces you to articulate what the project is about, and it ensures that the abstract exists from the start.

Every time you write a section, come back to the abstract and the introduction, keep working on them, and redo and polish them. Every time you write another section of the paper, come back to the abstract and revise it. Same for their spoken counterparts. Every time you present the project, notice which phrasings land and which fall flat, and update accordingly. By the end, hopefully, they are well done as you have spent time on them, without having to rush the night before submission.

As a reminder, it is tempting, especially in the final months before the job market, to keep adding robustness checks and extensions. Resist this. Of course, having a good paper with lots of extensions and robustness is important. The marginal return to a better-written, better-motivated, better-structured abstract is almost always higher than the marginal return to one more robustness check.

7.3 The Short Academic Product: Abstract and Elevator Pitch

The abstract and the elevator pitch are two expressions of the same product. They are **the most important piece of communication** you will produce. The abstract is what sits on the cover page of the paper; the elevator pitch is what you say when someone at a conference asks “So what are you working on?” when you meet a visiting seminar speaker, or in the opening minutes of a job-market interview. In each case, you have roughly one minute and the same elements to deliver.

The short product is your way through the door. It takes time, it takes practice, but because it is short and structured and always has the same elements, if you are disciplined in terms of knowing “this is roughly the words I have, these are the elements that I have”, it's feasible to do it and to do it well. If it fails to engage, the opportunity to communicate the research is lost. Once you are in,

you can spend time on the details. But again, you first need to get in.

7.3.1 Three Elements, One Limit

Good abstracts are surprisingly rare. Even candidates who can explain their project clearly in a two-minute conversation often have abstracts that are too long, poorly structured, and hard to read. There used to be a rule, imposed by the *Journal of Economic Literature* when it collected abstracts from other journals, that abstracts could not exceed **100 words**. Like the old character limit on social media platforms, this constraint was useful precisely because it forced researchers to distill their message to its essence. A practical approach is to draft a version of 150 to 200 words and then condense it.

A strong short product, whether written or spoken, contains **three elements**:

1. **Purpose:** a clear statement of the research question and its importance.
2. **Methodology:** a brief description of the methodology used to address the question.
3. **Findings:** a high-level summary of the main findings and why they matter.

Because the structures are parallel, preparing one well prepares you for the other. A tightly written 100-word abstract almost reads itself aloud as a pitch. Conversely, if you can explain your project clearly in 60 seconds of speech, you have the raw material for a strong abstract. If you need more words, the message is not yet sharp enough.

***Examples:** The best way to internalise what makes a good short product is to study well-crafted examples. The following four abstracts, drawn from different top journals, topics, and methods, all execute the three-element structure of a short academic product.*

1. **The Welfare Effects of Transportation Infrastructure Improvements**, Allen and Arkolakis (2022, REStud)

Abstract: Each year in the US, hundreds of billions of dollars are spent on transportation infrastructure, and billions of hours

are lost in traffic. We develop a quantitative general equilibrium spatial framework featuring endogenous transportation costs and traffic congestion and apply it to evaluate the welfare impact of transportation infrastructure improvements. Our approach yields analytical expressions for transportation costs between any two locations, the traffic along each link of the transportation network, and the equilibrium distribution of economic activity across the economy, each as a function of the underlying quality of infrastructure and the strength of traffic congestion. We characterize the properties of such an equilibrium and show how the framework can be combined with traffic data to evaluate the impact of improving any segment of the infrastructure network. Applying our framework to both the US highway network and the Seattle road network, we find highly variable returns to investment across different links in the respective transportation networks, highlighting the importance of well-targeted infrastructure investment.

The abstract opens with a vivid motivation: Each year in the U.S., hundreds of billions of dollars are spent on transportation infrastructure, and billions of hours are lost in traffic. One sentence, and the reader knows why this matters. Then the method: a quantitative general equilibrium spatial framework with endogenous transportation costs and congestion. Lastly, the finding: highly variable returns to investment across different links, which immediately implies that careful analysis matters for policy. Purpose, methodology, findings: all present, all clear.

2. **The Political Economics of Green Transitions**, Besley and Persson (2023, QJE)

Abstract: Reducing the emissions of greenhouse gases may be almost impossible without a green transition—a substantial transformation of consumption and production patterns. To study such transitions, we propose a dynamic model, which differs from the common approach in economics in two ways. First, consumption patterns reflect not just changing prices and taxes, but changing

values. Transitions of values and technologies create a dynamic complementarity that can help or hinder a green transition. Second, and unlike fictitious social planners, policy makers in democratic societies cannot commit to future policy paths, as they are subject to regular elections. We show that market failures and government failures can interact to prevent a welfare-increasing green transition from materializing or make an ongoing green transition too slow.

They first present the motivation/hook: Reducing the emissions of greenhouse gases may be almost impossible without a green transition, where they define the term explicitly, ensuring every reader follows. Then the method: “a dynamic model” differing from the common approach in two specific ways (changing values, not just prices and taxes; democratic policymakers, not social planners), with the contribution relative to existing literature is woven into it. Lastly, the finding: market failures and government failures can interact to prevent or slow a welfare-increasing green transition.

3. **Production Networks, Geography, and Firm Performance**, Bernard, Moxnes, and Saito (2019, JPE)

Abstract: This paper examines the importance of buyer-supplier relationships for firm performance. We develop a model in which firms outsource tasks and search for suppliers. Lower search and outsourcing costs lead firms to search more and find better suppliers, which in turn drives down marginal costs. We test the theory by exploiting the opening of a high-speed train line in Japan, which lowered the cost of passenger travel but left shipping costs unchanged. Using an exhaustive data set on firms’ buyer-seller linkages, we find significant improvements in firm performance as well as the creation of buyer-seller links, consistent with the model.

The first sentence states the paper’s purpose: This paper examines the importance of buyers and suppliers for firm performance. Again, one sen-

tence, and you know exactly what the paper is about. Then it presents the method: combining a model of outsourcing and supplier search with a natural experiment, the opening of a high-speed train in Japan. Lastly, the finding: significant improvements in firm performance as well as the creation of buyer-seller links, consistent with the model.

4. **What Drives Media Slant? Evidence from U.S. Daily Newspapers,** Gentzkow and Shapiro (2010, ECTA)

Abstract: We construct a new index of media slant that measures the similarity of a news outlet’s language to that of a congressional Republican or Democrat. We estimate a model of newspaper demand that incorporates slant explicitly, estimate the slant that would be chosen if newspapers independently maximized their own profits, and compare these profit-maximizing points with firms’ actual choices. We find that readers have an economically significant preference for like-minded news. Firms respond strongly to consumer preferences, which account for roughly 20 percent of the variation in measured slant in our sample. By contrast, the identity of a newspaper’s owner explains far less of the variation in slant.

Here, the novel measurement strategy is explained up front: an index comparing a news outlet’s language to that of congressional Republicans or Democrats. Then: a demand model that estimates profit-maximising slant. The key result is immediately memorable: consumer preferences account for over 20% of variation in slant, while owner identity explains far less. The abstract is short, yet the research question, the method, and the findings are all unambiguous.

Across all four examples, you see a very clear format. The elements are always the same. First, some motivation and a question. Second, since you’re going to end up talking about what you find, it cannot be just pulling a rabbit out of a hat, but needs to be a research output that people can trust. So you need to give a hint about the methodology. And then finally, the findings, then the actual core

results. Sometimes they give you a number, often none. So it’s a very high-level summary of the findings. Typically, it also contains a hint of why the research is important.

7.4 The Long Academic Product: Introduction and Spiel

The long product extends the short product into a richer, more structured argument. For the written format, this would be the introduction of a paper, which is the most important and most difficult section to write. In terms of oral communication, the opening minutes of a seminar serve the same function. Besides, in a job-market interview, where you typically enter a Zoom meeting with a panel, you will be asked to “tell us about your research”. The first bit of the job market spiel is also going to be similar.

7.4.1 Canonical Templates and Essential Elements

Below are two well-known frameworks for structuring an introduction, both capturing the essential ingredients. Essentially, they have the same elements: attracting the reader’s interest, stating the research question clearly, situating the paper within prior work, explaining its contribution, and sometimes outlining the paper’s structure.

Table 2: Two Canonical Formulae

Element	Keith Head’s Formula	Don Cox’s “Big Five”
1	Hook	Tell your research question immediately.
2	Question	Justify its importance.
3	Antecedents	Spell out the deficiencies in previous work.
4	Value Added	Explain what you are doing that’s better.
5	Roadmap	Summarize your results right up front.

1) Hook and Question There are two ways to structure the first part of the introduction. One way to go is to first give the hook, something that provides motivation, and then go to the specific. The second possibility is to start from the very specific, your research question, and then go broader and explain why it matters. Both work.

A hook is used to grab people’s attention by telling them that the paper tackles something important. It is important to keep it short. But be aware that a very broad motivation (“inequality is important”) is not a hook or motivation; a narrow, vivid fact tied directly to your project is. Some researchers seem to be worried about revealing too much too early, but remember, you are not writing a mystery novel!

For the question, people need to know explicitly what the paper is about in the first paragraph of the paper. A good way to write this is to start with one sentence and then go into more detail, because formulating the whole question should be longer, maybe one or two paragraphs. It’s important to separate the question from the answer, so make sure you fully state your question before you actually start answering it.

2) Answer: method, results, and takeaway After fully stating the question, you need to talk about your answer and how you reached it. Therefore, you need to build trust by explaining the method: the type of model, the data, the key source of variation, and the identification strategy. Then state the results, identifying **one key takeaway** the reader can remember. When doing so, do not be afraid of committing to a clear message. There will always be caveats, but the core takeaway should be unmistakable. Restrict yourself to one or two numbers; more detail belongs in the body. Aim to keep the full introduction to **roughly two pages**.

3) Literature and Contribution The third part of the introduction is devoted to positioning your paper in the literature. Discussion of prior work should be selective and purposeful. The goal is not to cite everything, but to establish the foundations for the paper’s contribution. Every reference should answer a question of the reader: *What do I need to know about prior work so that I understand what*

is new about yours? References that do not serve this purpose can be mentioned later in the paper.

You need to explain what is important about what people have done before, give credit where credit is due, but also make clear what the shortcomings are, especially shortcomings that you are going to address. If there is a closely related or similar paper that you need to differentiate yours from, the best person to do it is yourself. Do it explicitly as opposed to hiding it.

Literature review and the contribution part often come together. A way to state contribution while giving credit to prior work could be: “They did this, which is very important for this reason. However, they had this shortcoming, and my research is new because I solved this problem”. Note that the contribution is different from the main findings. The findings are the results. Implicitly or explicitly, this is the place where you should give an argument for why the paper should be published.

4) Limitations and Roadmap No paper is the final word on something; there’s always going to be shortcomings. If there are shortcomings you cannot fully resolve, it is often better to acknowledge them in the introduction and explain why they should not be overly worrying, rather than simply ignoring them. You know the paper better than anybody; you can find the argument better. But keep the focus on the main issues, as the introduction is not the place to defend against every possible criticism.

Typically, you finish the introduction with a roadmap. However, if the roadmap is very standard, such as “data, methodology, results, conclusions”, then you might skip it. If you have already described what you do very linearly, you can also skip it to avoid redundancy.

Once you have a complete draft of the long product, perform a systematic **coherence check**. A well-written paper forms a coherent whole. The research question, the answer, and the discussion of prior work should reinforce one another. Useful consistency checks include asking yourself some questions: Does the motivation make readers care about your *specific* question? Does the answer really read as an answer to *that* question, or to a subtly different one? If a reader

saw only the shortcomings you identify in prior work, would they already guess what your paper does? Does the description of your contribution match the gaps you outlined? Making sure all the pieces fit together.

7.4.2 Bonus: Rest of the Paper

The long written product naturally extends beyond the introduction into the body and conclusion of the paper. While the abstract and introduction are the most difficult and highest-return sections to write, the rest of the manuscript also constitutes part of the research product and should be crafted accordingly.

Paper structure Standard structures exist for empirical and theoretical papers. If you deviate, have a good reason, because readers will expect the convention.

- **Applied empirical paper:** Introduction → Background → Model/Empirical Strategy → Data → Results (possibly split into multiple sections) → Conclusion.
- **Applied theory paper:** Introduction → Motivating Evidence → Model → Results → Extensions → Evidence Supporting Model Implications → Conclusion. This “theory sandwich” is typical: motivating evidence informs the model, and the model generates additional predictions (difficult to come up with without the model) that are then tested empirically.

In addition, in both oral and written forms, the **pyramidal structure** applies: bottom line up front, then details developing the argument, and finally background and supplemental information. In this way, you go from less to more. Notice that when we discussed both the short and long products, we already had this structure: bottom line up front (first state research question), details to develop the argument (answer), and then antecedents/contribution/roadmap are the background and context. This principle scales down to individual sections and even paragraphs: each paragraph should open with a topic sentence that previews its content. A useful exercise for the written products: take the first sentence of each paragraph in your introduction, read them in sequence, and check whether they read like a slightly longer abstract.

Tables and figures Tables and figures are great for conveying results. They should be **self-contained**: clear titles, labeled axes, explicit variable names, estimation methods noted, and footnotes providing enough details that a reader can understand the exhibit without consulting the main text. Many readers browse a paper by reading the abstract, then the introduction, then skimming the tables and figures. If those elements do not stand alone, you lose that reader.

Adapt tables and figures for different contexts. Tables in seminar slides should focus on the key variables. For figures, label curves directly rather than relying on external legends, and choose color-and-pattern combinations that remain legible in black and white.¹⁴

Theory Any model is a representation of reality with enough simplification to distill clear conclusions yet not so much that it becomes unrealistic or irrelevant. Therefore, you need a careful **balance between simplicity and realism** for your model. When you have a model, there will be many assumptions. Solow’s famous growth paper starts with “All theory depends on assumptions which are not quite true. That is what makes it a theory. The art of successful theorizing is to make the inevitable simplifying assumptions in such a way that the final results are not very sensitive. A ‘crucial’ assumption is one on which the conclusions do depend sensitively, and it is important that crucial assumptions be reasonably realistic.” Therefore, for every assumption, do two things: explain *why* it is there and what feature of reality it captures, and state whether it is made for analytical convenience or whether it is *crucial* for the results. If an assumption is crucial, provide evidence that it is realistic.

Conclusions Keep conclusions short, typically one page. This is where you summarize key takeaways and limitations, but also somewhere to have the liberty to be more speculative than in the rest of the paper. You can discuss policy implications, suggest directions for future research, or sketch how the project could be extended. While speculation is acceptable here, it still should be grounded in your research. Krugman discussed in his 1980 AER paper that allowing factor

¹⁴Lecture 3 provides a comprehensive treatment of data visualization and the visualization of econometric results.

mobility might produce agglomeration, an idea that became the foundation of the New Economic Geography eleven years later. Of course, your speculations in the conclusion do not need to go this far, but it is still a good place to say that the work is part of a broader agenda, or to preempt criticism by highlighting how you could have tackled things that might worry people, which you are not doing but would reasonably not go against your research.

Lastly, as an additional brief comment for producing research outputs. Nowadays, new tools such as artificial intelligence can assist with specific tasks, such as editing or improving clarity (and are super good for coding). A recommendable way of using AI for writing is to guide it carefully and use it to refine existing text, or when there is already a satisfying initial structure. Think of AI as a capable assistant that helps you think, code, and write more efficiently, not as a replacement for your own thinking and writing.

7.5 Policy-Oriented Research Products

Academic research products are not the only outputs an economist can produce. Policy-oriented research constitutes a parallel category with its own formats and constraints. Understanding these formats is valuable even at the PhD stage. Policymakers may seek research evidence to inform decisions, to justify actions already taken, or to improve implementation. There are roughly three types of policy-related research outputs:

1. **Policy Evaluation:** Academic journals prioritise novelty, but policy-oriented research often requires answering a question for a concrete measure important for the decision. For instance, evaluating a transport program, a labor-market reform, or a tax change for a particular setting is a genuine input into the decision process. Therefore, even if they probably will not yield a top-journal publication, these papers are still valuable products.
2. **Advisory Input:** Policymakers sometimes seek insights from researchers to inform decisions they have not yet made. This is probably going to happen later on in your career, such as taking the role of an Economic Advisor. For

this, policymakers would want to talk to you and obtain information that you know about because you are a researcher. Therefore, producing this format of output involves identifying research informative for their decisions, and translating academic findings into accessible and actionable knowledge.

3. **Policy Justification:** When policymakers have already decided on doing something, they will want to seek some justification and hope that you can provide some research to support it. This is more common than one might expect, but it varies greatly across countries.

The boundary between academic and policy research is evolving. As data-driven analysis and more sophisticated exercises become more common within governments, the demand for economists who can do rigorous research in policy settings is growing¹⁵. This means that policies are much better grounded in solid research, but it also has implications for your job-seeking process. For PhD students, the skills you develop, such as careful identification, transparent analysis, and clear communication, have value well beyond the walls of academia, but it also requires greater flexibility and adaptability of these skills in the policy domain.

¹⁵In Spain, for example, institutions such as AIREF and the Bank of Spain regularly recruit trained economists. Every ministry now has a scientific advisor serving as an interface between the academic community and the minister, and a recent open call invites researchers to work embedded within ministries for six to eight months on specific policy questions.

Afterword

We never thought this would end up being such a public good. But here we are!

The *BESTEPS Guide* grew out of the BESTEPS Lectures, a flagship PhD training project organized at the UB School of Economics (UBSE). But it is more than just a collection of lectures.

Just as generating a research idea begins with identifying a relevant and answerable gap in our knowledge, designing the BESTEPS project followed the exact same logic. For us, the gap was a set of research skills missing from the standard PhD curriculum, skills that are not only practical but also sharpen the judgement and research sense that are more important than ever in the age of AI. Looking back, we are glad that we took up the opportunity and built a fully demand-driven curriculum to fill it.

Such a curriculum, taught by leading experts, is too valuable to keep to ourselves. This *Guide* is our attempt to turn the lectures into a shareable resource, one that can lower the barriers for junior researchers around the world facing the same steep learning curves. It was not easy: we invested more than 100 hours turning it into what we hope is an enjoyable read.

A project like BESTEPS does not appear in a vacuum. It needs a community that is at once rigorous and collaborative, and that takes early-career researchers seriously. Years of collective effort have made UBSE exactly such a place, with its PhD program playing a central role. Here, faculty and students work in close contact, junior researchers bring early-stage ideas to the biweekly “Brainstorm Coffees”, and PhD students organize seminars across universities in Barcelona. This is the UBSE that made BESTEPS possible.

BESTEPS is, ultimately, an expression of this broader vision: we have seen what a research community can give to those just starting out, and we believe what it gives should not stay behind our own doors. This *Guide* is our small way of contributing to the wider academic community we belong to. The research journey is hard for all of us; wherever you are along the way, we hope the *Guide* meets you there.

Mengwei Lin

BESTEPS Coordinator & PhD Candidate
University of Barcelona & IEB

Jordi Roca Solanelles

Head Manager, UBSE
University of Barcelona