Universidad de los Andes
Facultad de Economía

ECONOMIA URBANA

2025-2
Econ 4683

Ignacio Sarmiento-Barbieri

Final Project Proposal

The goal of the final project is to define a research question and to develop a causal research design in urban economics. This design should be grounded in the literature and supported by **simulated data** that enables a full design diagnosis.

A good proposal:

- Begins with a well-defined topic.
- Motivates and argues for the importance of the topic.
- Demonstrates that the author is well acquainted with the relevant literature.
- Outlines a compelling methodological approach.

This is a *design project*. You will not collect real data. Instead, you will **simulate plausible data** based on your proposed research design and use it to explore how well your strategy can recover the effect of interest.

Your project should consist of:

- A document with research proposal (written like a paper),
- A script that contains your simulations and analysis.

Your proposal should follow the template below, each subsection gives guidelines on what should be in that section. See the additional guidelines at the end for more details.

Universidad de los Andes
Facultad de Economía

ECONOMIA URBANA

2025-2
Econ 4683

Ignacio Sarmiento-Barbieri

(Insert study title here)

Authors: Provide name of all authors

Abstract

Limit the abstract content to a maximum of 150 words. An abstract is often presented separately from the article, so it must be able to stand alone. Use the following questions to guide you:

- What is/are the research question(s) of this study and why is it important for the urban economics literature?
- What is the main outcome variable; what is/are the secondary outcome variable(s) (if any) and what is the research question?
- What kind of methodological framework will the study use to address the research question?

Keywords: Provide up to four keywords in for indexing purposes.
JEL codes: Provide up to four codes from those available here.

Universidad de los Andes
Facultad de Economía

ECONOMIA URBANA

2025-2
Econ 4683

Ignacio Sarmiento-Barbieri

1 Introduction

This section addresses the proposal's research question: background, importance and relevance. Write this section following *Keith Head's* formula. To help you write this section keep in mind these guiding questions:

- What is the main problem/question motivating the study? Why is this question important for the field of urban economics?
- How has this question been addressed thus far in the relevant literature? What are the competing theories for explanation of this question? How is this study different from prior research on this problem/question?

Moreover, the proposal should contain at least ten relevant references when reviewing the relevant literature. At least two of those references must be from a **top ten journal** publishing original research. A top ten journal in economics is hard to define. However, we will use the h5-index rankings since it produces a good approximation. For example, on October 2023, these were: American Economic Review, The Review of Financial Studies, Journal of Political Economy, The Quarterly Journal of Economics, The Journal of Finance, The Review of Economic Studies, Econometrica, Journal of Public Economics, the Review of Economics and Statistics and the Journal of Development Economics.[1]

2 Research Design

This section is the core of your proposal. It should clearly articulate your causal inference strategy and convince the reader that your approach can credibly identify the effect of interest. A strong research design section demonstrates both theoretical rigor and practical feasibility.

2.1 Causal Framework

Begin by establishing the causal framework for your study:

- Define the causal estimand: What is the precise causal quantity you aim to estimate? (e.g., ATE, ATT, CATE, LATE, ITT). Be specific about the population and the treatment contrast.
- State the potential outcomes: Using the potential outcomes framework, clearly define $Y_i(1)$ and $Y_i(0)$. What would happen to unit i under treatment versus control?

[1] I've excluded the Journal of Economic Perspectives since its purpose is not to publish novel research but to *"bridge the gap between the general interest business and financial press and standard academic journals of economics"*

- Articulate the identifying assumption: What key assumption allows you to identify your causal estimand? (e.g., random assignment, parallel trends, exclusion restriction, continuity at the threshold). Explain this assumption in plain language and discuss its plausibility in your context.

2.2 Identification Strategy

Describe your identification strategy in detail:

- Name and justify the approach: Are you using an RCT, difference-in-differences, instrumental variables, regression discontinuity, or another strategy? Why is this approach appropriate for your research question?

- Explain the source of variation: What generates the variation in treatment? Is it a policy change, a randomized intervention, a threshold rule, or natural variation? Be specific about the institutional details.

- Discuss threats to identification: What could go wrong? Consider threats such as:

  - Selection bias or endogeneity
  - Spillovers or general equilibrium effects
  - Attrition or non-compliance
  - Measurement error
  - Pre-existing trends (for DID)
  - Manipulation of the running variable (for RD)

  For each relevant threat, explain how your design addresses it or why it is unlikely to be a concern.

- External validity: To what population or setting do your results generalize? What are the limitations of your design in terms of external validity?

2.3 Estimation Strategy

Provide the formal econometric specification:

- Baseline specification: Write out your main regression equation. For example:

$$Y_{it} = \alpha + \beta D_{it} + \gamma X_{it} + \delta_i + \theta_t + \varepsilon_{it} \tag{1}$$

where D_{it} is your treatment variable, X_{it} are controls, δ_i are unit fixed effects, and θ_t are time fixed effects.

- Interpret the coefficient of interest: What does β represent in your context? What are its units?
- Specify standard errors: How will you calculate standard errors? Will you cluster? At what level? Why?
- Additional specifications: If relevant, describe:
  - First-stage and reduced-form equations (for IV)
  - Event study specifications (for DID)
  - Local polynomial regressions (for RD)
  - Heterogeneity analysis (interactions or subgroup analysis)
- Sample restrictions: Will you impose any sample restrictions? (e.g., bandwidth choice for RD, common support for matching). Justify these choices.

2.4 Robustness and Specification Checks

Outline the robustness checks you plan to conduct:

- What alternative specifications will you run?
- Will you test pre-trends (for DID), check for bunching (for RD), or validate the instrument (for IV)?
- How will you assess the sensitivity of your results to key design choices?

These checks should be described here conceptually, even though the actual results will appear in the Results section.

3 Data

This section describes the simulated data you will use to evaluate your research design. While these data are synthetic, they should be constructed to realistically reflect the empirical setting you propose to study. Your goal is to demonstrate that you understand the data-generating process relevant to your question and can design a simulation that captures the key features of real-world data.

Write this section as if you were describing real data in a published paper and provide sufficient detail that another researcher could reproduce your simulation.

3.1 Data-Generating Process

Describe the process you used to generate the data:

- Units of observation: What are the units in your data? (e.g., individuals, households, neighborhoods, cities, firms). How many units do you simulate?
- Time structure: Is your data cross-sectional, panel, or repeated cross-sections? If panel, how many time periods?
- Treatment assignment mechanism: How is treatment assigned in your simulation?
  - For RCTs: describe the randomization procedure (simple, blocked, clustered)
  - For quasi-experiments: describe the rule or process that determines treatment (e.g., threshold for RD, policy timing for DID)
- Outcome variable: How do you generate the outcome variable Y? Specify:
  - The functional form (e.g., linear, log-linear)
  - The true treatment effect you embed in the data
  - The role of covariates
  - The error term distribution and properties
- Covariates: What covariates do you include? How are they generated? Are they:
  - Correlated with treatment? (If so, how?)
  - Correlated with the outcome? (Specify the mechanism)
  - Measured with error?
- Fixed effects (if applicable): Do you include unit or time fixed effects? How do you generate them? Are they correlated with treatment assignment?
- Error structure: Describe the properties of the error term:
  - Distribution (normal, skewed, heavy-tailed?)
  - Homoskedastic or heteroskedastic?
  - Clustered? (intra-cluster correlation coefficient?)

  - Serial correlation? (if panel data)

- Complications (if any): Do you include:
  - Non-compliance or treatment spillovers?
  - Attrition or missing data?
  - Measurement error?
  - Endogenous selection?

  If so, describe how you model these features.

Justification: For each design choice, briefly explain why it is realistic or relevant to your empirical context. For instance: ”We simulate heteroskedastic errors because housing prices exhibit greater variance in high-income neighborhoods” or ”We include a 15% non-compliance rate consistent with findings from prior field experiments.”

3.2 Descriptive Statistics

Present descriptive statistics that characterize your simulated data. This subsection should mirror what you would include in an empirical paper using real data.

- Summary statistics table: Include a table showing means, standard deviations, min/max for key variables, separately for treatment and control groups (if applicable).
- Balance checks (for experimental designs):
  - Present a balance table comparing treatment and control groups on pre-treatment covariates
  - Include t-tests or F-tests for joint significance
  - For block or stratified designs, show balance within blocks
- Pre-trends visualization (for DID):
  - Plot trends in outcomes for treated and control units in pre-treatment periods
  - Discuss whether parallel trends appear satisfied
- Discontinuity visualization (for RD):
  - Plot the outcome variable against the running variable
  - Show the discontinuity at the threshold

- Test for manipulation of the running variable (e.g., McCrary test)

- First-stage relationship (for IV):
  - Show the relationship between the instrument and the endogenous variable
  - Report the F-statistic for instrument strength

- Geographic or spatial patterns (if relevant):
  - Include maps showing the spatial distribution of treatment or key variables
  - Discuss clustering or spatial autocorrelation if relevant

- Other relevant descriptives: Use your judgment about what descriptive statistics best tell the story of your data and research question.

4 Results

Present and interpret your main findings:

- Estimation results: Present your main regression results in a well-formatted table. Include:
  - Point estimates with standard errors
  - Statistical significance indicators
  - Sample size and R-squared
  - Multiple specifications if relevant (e.g., with/without fixed effects, with/without controls)

- Interpretation: Discuss the magnitude, sign, and statistical significance of your estimates. What is the economic significance of the effect? Put the magnitude in context.

- Specification comparisons: Compare results across specifications to demonstrate the importance of your identification strategy. For example:
  - If fixed effects are part of your identification, show results with and without them
  - If using block randomization, show what happens with and without accounting for blocks
  - If using DID, show the naive cross-sectional comparison vs. the DID estimate

- Robustness checks: Present results from the robustness checks outlined in your Research Design section.

- Theoretical interpretation: Interpret your results in light of competing theories in urban economics. Which theories do your findings support or reject?

- Policy implications: What are the potential policy implications of your findings?

5 Conclusion

This section should briefly synthesize your main contributions and provide a roadmap for implementation:

- Summary of findings: Restate your main research question and summarize your key findings. What did your simulated design reveal about the effect of interest?

- Contributions to the literature: How does your proposed design advance our understanding of the research question? What does it add to the urban economics literature beyond existing studies?

- Design insights: What did you learn from the design diagnosis exercise? What are the key trade-offs in your proposed design? What design parameters are most critical for obtaining credible estimates?

- From simulation to implementation: Reflect on the transition from simulated to real data:
  - Data requirements: What specific data would you need to implement this design? Where could you obtain it? (government agencies, administrative records, surveys, web scraping, partnerships with firms/municipalities?)
  - Timeline and resources: What would be a realistic timeline for data collection? What financial and human resources would be required?
  - Institutional partnerships: Would you need to partner with government agencies, NGOs, or private firms? What would be the challenges in establishing these partnerships?
  - Ethical considerations: What ethical approvals would be needed? Are there privacy concerns or potential harms to consider?
  - Unexpected challenges: What complications might arise in real implementation that your simulation did not capture? (e.g., non-compliance, attrition, measurement error, political constraints)

- Deviations from the ideal design: Real-world constraints often force researchers to deviate from ideal designs. What compromises might you need to make? How would these affect your ability to identify the causal effect?

- Limitations and extensions: Acknowledge the limitations of your design. What simplifying assumptions did you make? What aspects of the problem does your design not address? What would be valuable directions for future research?

- Final remarks: Close with a brief statement about the importance of the research question and the next steps you would take to move from design to implementation.

6 References

Include all the references cited in the text.

Universidad de los Andes
Facultad de Economía

ECONOMIA URBANA

2025-2
Econ 4683

Ignacio Sarmiento-Barbieri

Additional Guidelines

Deliverables

Your final submission should include both a written report and the code used to generate your analysis. This project is not only about designing a compelling empirical strategy but also about communicating it clearly and ensuring full reproducibility.

The written report must be submitted as a single PDF file on Bloque Neón. The document should be no longer than **15 (fifteen) pages** and may include up to **10 (ten) exhibits** (tables and/or figures). Use tables and figures effectively. Each exhibit should be clearly labeled, include notes explaining variables and samples, and be referenced in the text. Think carefully about what information is most important to convey, quality matters more than quantity. Bibliography and exhibits do not count toward the page limit. You are welcome to add an appendix, but the core document should be self-contained and coherent.

The document should follow these formatting guidelines:

- A4 paper size
- 1.5 line spacing
- 2.5 cm margins
- 11pt font size
- Use the default font in LaTeX (Computer Modern) or `Times New Roman` if using Microsoft Word

In addition to the PDF, you must submit a compressed folder that contains:

- **The full code used to run your simulations and analysis**. It must be fully reproducible and produce all results shown in the report.
- **A `README` file** that documents the file structure and provides clear instructions on how to run your code.
- **Readable and well-commented code**: Good coding style is essential. I recommend following the tidyverse style guide.

Group work. You may work in teams of up to **three** members. These can be the same groups you worked with during the semester or new ones. Students in the PEG or Ph.D. programs are **strongly encouraged to work independently** and use this project to develop material that could feed into their thesis. *Please notify me by **November 21** if you are changing groups or working solo.*

Universidad de los Andes
Facultad de Economía

ECONOMIA URBANA

2025-2
Econ 4683

Ignacio Sarmiento-Barbieri

Evaluation Criteria

Your final grade will be based on several dimensions:

Introduction (20%)

A strong introduction demonstrates mastery of the literature and clear motivation for the research question.

- **Quality and appropriateness of references**:
  - Does the proposal cite at least **10 relevant references**?
  - Are at least **two references from top journals**?
  - Are the references recent and directly relevant to the research question?
  - Does the literature review go beyond summarizing papers to synthesize competing theories and identify gaps?
- **Following Keith Head's formula**:
  - **Hook**: Does the opening motivate why the question matters for urban economics?
  - **Question**: Is the research question clearly stated early on?
  - **Antecedents**: Does the proposal review what we know and what remains unresolved?
  - **Value-added**: Is it clear how this study differs from and builds upon prior work?
  - **Road map**: Does the introduction preview the empirical strategy and main findings?
- **Conceptual clarity**: Are the economic mechanisms and theoretical predictions clearly articulated?

Complexity of the Design (25%)

More sophisticated identification strategies will receive higher grades. A rough progression could be:

Level	**Design Types**
Basic	Audit experiments, List experiments, 2-arm RCTs
Intermediate	Block randomization, Cluster randomization, Selection-on-observables with fixed effects, Difference-in-differences
Advanced	ITT/LATE analysis, Staggered Difference-in-differences, Instrumental Variables, Regression Discontinuity

Note: Complexity alone does not guarantee a high grade. A well-executed intermediate design is better than a poorly justified advanced one. The key is to match the design to the research question and to demonstrate deep understanding of the identification strategy.

Realism of the Simulated Data-Generating Process (20%)

- Is the simulation coherent with the proposed theory and empirical context?
- Are fixed effects, error terms, and covariates modeled plausibly?
- Does the simulation reflect real-world features such as noise, measurement error, selection, or non-compliance?
- Are the parameter values (e.g., treatment effects, variance components) realistic and justified?
- Does the descriptive analysis effectively characterize the simulated data?

Realism of the Design (15%)

How plausible is the proposed design in the context of urban policy or research?

- Could this research design be implemented in a real-world setting (e.g., a city, municipality, or housing market)?
- Are the identifying assumptions realistic given typical data limitations and institutional constraints?
- Are the outcome variables, units of observation, and treatment definitions credible?
- Does the "From simulation to implementation" discussion show awareness of practical challenges?

Writing and Presentation (20%)

Clear communication is essential in research. Your document will be evaluated on:

- **Adherence to template**: Does the proposal follow the required structure? Are all sections present and substantive?
- **Academic writing style**:
  - Is the writing clear, precise, and professional?
  - Are technical terms defined and used correctly?
  - Does the writing flow logically from section to section?
  - Is the tone appropriate for an academic economics paper?
- **Tables and figures**:
  - Are exhibits professional, clearly labeled, and easy to interpret?
  - Do tables include appropriate notes defining variables, samples, and significance levels?
  - Are figures referenced and discussed in the text?
  - Is the number of exhibits reasonable given the exhibit limit?
  - **Follow top-journal conventions:** Before submission, open two recent papers you cite and mirror their conventions for titles, captions, notes, units, and numbering. Keep our style consistent with those examples.[2]
- **Reproducibility and code quality**:
  - Is the code well-organized, commented, and reproducible?
  - Does the README file provide clear instructions?
  - Does the code follow good practices (e.g., tidyverse style guide)?
- **Grammar and formatting**: Is the document free of typos and grammatical errors? Does it follow the specified formatting guidelines (margins, spacing, font)?
- **Conciseness**: Does the proposal respect the page limit while conveying all essential information? Quality over quantity.

[2] **Pre-submission check:** Pick one table and one figure, compare against those papers, verify title, caption or notes, units, uncertainty display, and cross-references.

Universidad de los Andes
Facultad de Economía

ECONOMIA URBANA

2025-2
Econ 4683

Ignacio Sarmiento-Barbieri

Bonus (up to 10% of the grade)

You can earn up to an additional **10% bonus** if your project meets the following criteria:

1. **Conciseness**: Projects that are notably clear and concise may be awarded bonus points. As Hamlet put it: *"Brevity is the soul of wit."*

2. **GitHub Repository**: Instead of submitting a compressed folder, you provide a link to public GitHub repository where your files are hosted. The repository should include:

   - A clear and helpful `README` file that guides the reader through your project. This file should be informative enough to explain what the project does and how to reproduce the results, but concise enough to keep the reader engaged. For examples of strong READMEs, see the curated list at Project Awesome.
   - Well-structured folders and code files that mirror your analytical workflow.
   - Visible contributions from all group members in the commit history.

You do **not** need to meet all of them to receive bonus points; partial bonuses may be awarded.