Senior Thesis

Economical writing

Kyle Coombs (adapted from work by Marc Bellemare and Keith Head)
Bates College | EC/DCS 456

Table of Contents

Prologue

Prologue

- Today we're going over writing an economics paper

- We're going to talk about how to write a paper for the audience

  - Hint: they're not reading it for the worldbuilding or the plot twists like a novel

- Then we'll talk through how to structure the sections

- And finally we'll talk about how to write the sections

  - Note that introduction comes last because it's easier to write when you know what you're introducing
  - But you'll be writing it first, this is just to say that you should go back and redo it later!

Inspectional reading

- Academics do not read every word of a paper

- They inspect the paper to see if it's worth reading further and stop when it isn't

- That means you should assume your reader wants to read something else at all times

- You have to hook them early and keep them hooked throughout

- There are sections of your paper that they will skip entirely

    - Unless they want to check your work, try to replicate it, or dig into it

- So get to the point and keep it interesting

Which parts are least read?

- Appendix?

- Theoretical model?

- Data and descriptive statistics?

- Introduction?

- Abstract?

- Conclusion?

Structure

Structure

1. Title

2. Abstract

3. Introduction

4. .....

5. Summary and Conclusion

6. References

- Wait, what goes in 4?

- Depends on the paper

Applied paper structure

1. Title
2. Abstract
3. Introduction
4. Theoretical Framework
5. Data and Descriptive Statistics
6. Empirical Framework
7. Results and Discussion
8. Summary and Concluding Remarks
9. References
10. Appendix

- Notice anything odd?

Structure is not set in stone

- Stay flexible -- your paper might need a full lit review

- Maybe you want to add a section explaining the institutional setting in case the reader is unfamiliar

- Maybe there's little theoretical component, so that gets wrapped up in the introduction/empirical framework

- Maybe the only empirical work is for a robustness check, so you leave that for the Discussion

- But learn the rules before you break them

Theoretical Framework

Theoretical framework

- Most applied economics answers: "Does x cause y" or "If x changes by 1, by how many units does y change?"

- So most applied economics papers have some theoretical framework of change

- Remember the causal question of interest from the FAQs?

  - You might be asking how an additional year of schooling affects wages, but you're also asking how additional human capital increases wages

- A theoretical framework usually has a model that simplifies the world to provide a hypothesis about how the world works

Inspectional readers

- Inspectional readers are likely not reading into every lemma of every proof

- They're reading to work out the heuristics of the model and the intuition of the results

- Any well-written applied theoretical model should be easily explained in a few sentences or a toy example

Classic theoretical framework

- Akerlof (1970) "The Market for Lemons"

- Adverse selection in markets with asymmetric information: some cars are "lemons," some are not

- Car sellers know a car's quality

- Buyers don't know and infer quality from price: p^H for high quality, p^L for lemons

- Buyer assumes some probability of getting a lemon, so is willing to pay less than p^H

- Sellers unwilling to sell at p^L, so only lemons are sold and the market can unravel

- The paper uses a simple model in which utility is a function of the price paid and the quality of the car

- The model predicts an information intervention (e.g. a warranty) can restore the market

- An empirical paper on random assignment of warranties would be a test of the model

Theoretical framework

- Best question to ask:

1. Am I working on something theorists have worked on before? (All of you are.)

2. Am I working on something theorists have not worked on before?

- If 1, then you can use the existing literature to guide your model

    - Maybe you can take it off the shelf

- If 2, then you have to build your own theory

    - This may lead to a very different paper

Data and descriptive statistics

Data and descriptive statistics

- An inspectional reader uses this section for fast facts about the data

- They want to know the external validity and internal validity of the data

- That means they need to know three things:

1. Metadata

2. Variable definitions

3. Critical summary statistics

Metadata

- There's lots of basic information about your data that a new person to your research question won't know

  - Where is the data from?
  - When was it collected?
  - By whom was it collected?
  - What is the unit of observation? sample size?
  - How was it collected? (A random sample, a census, digital records, etc.)
  - What's the ideal data (sample size? units of observation? etc.)?
  - Other missing data? How'd you deal with that? Imputation?
  - Any nonresponse? Any attrition with time? (is it related to treatment!)

- Basically, the reader wants to know the external validity and the internal validity of the data

Variable definitions

- This is in many ways the most tedious part of the paper

- It is where you precisely define all the variables you use in the paper

- For example, are you defining unemployment as any person who is not employed?

    - Or anyone with an observed job separation?
    - Or anyone with an observed firing from a job?

- Sometimes this will get pushed into an appendix upon publication but it's still important to have it in the paper

- And sometimes it is a great to put it all in a table for quick reference

A nice table

Data Description for Selected Variables

Variable	Description
Dependency ratio	Percentage of individuals under 15 and over 64 within the household
Assets (100,000 ariary)	Sum of the values of the household's assets (i.e., animals, house, television, radio, car, and bank account balance) and agricultural equipment (i.e., hoe, harrow, cart, plow, tractor, and small tractor)
Income (100,000 ariary)	Sum of the proceeds from animal sales, agricultural and nonagricultural wages, and proceeds from leases of cattle and equipment
Liquidity constraint dummy	Dummy for whether the household is liquidity constrained
Plot size	Area covered by the plot in ares (1 are $=$ 001 ha $=$ 100 m^2)
Plot value (100,000 ariary)	Price expected by the landowner if she were to sell her plot
Formal title dummy	Dummy for the presence of a formal title
Relationship length	Number of years the landlord and tenant have been contracting with one another
Kin dummy	Dummy for a contract signed between kin
Tenant introduced by kin	Dummy for a contract signed with a tenant whom the landlord met through a member of her extended family
Introduced by other than kin	Dummy for a contract signed with a tenant whom the landlord met through someone who is not a member of her extended family
Tenant is friend	Dummy for a contract signed with a tenant who is a friend of the landlord
Tenant chosen for his wealth	Dummy for whether this particular tenant was chosen because of his wealth
Tenant chosen for his honesty	Dummy for whether this particular tenant was chosen because of his honesty
Tenant chosen for his ability to bear risk	Dummy for whether this particular tenant was chosen because of his ability to bear risk
Tenant chosen to return a favor	Dummy for whether this particular tenant was chosen because the landlord wanted to return a favor
Time spent looking for a tenant	Number of days spent looking for a potential tenant
Other potential tenants considered	Number of other potential tenants considered when looking for a tenant

Summary statistics

- Summary statistics are the descriptive statistics of your data

- There are tons of ways to cut data, so make sure you show the most important:

  - Is there a treatment/control framework? Show the means of the treatment and control groups for many variables and a t-test of the difference (balance table)
  - Is the main effect of interest visible in one chart like the average stock return after an FOMC announcement? Show that chart!
  - Maybe add some density plots (continuous) and histograms/cross-tabs (categorical) to show the distribution of key outcomes (show the reader if your results are driven by outliers or not)

- Common mistakes to avoid:

  - Don't write up every random average or statistic you can grab in the data even for obscure controls
  - Don't speak in the past tense (keep it active!)
  - Don't add tons of decimal places or odd scientific notation -- instead rescale variables to be more interpretable

Ideal Data

- Early on, I think it is great to have an ideal data section!

- This is to help you wrap your head around what your data's limitations are so you can better respond to your reader

- For example, are you measuring income the ideal way?

- How far off is your sample selection from the ideal random sample?

- Eventually, just wrap that up in the rest of it all

Empirical Framework

Empirical Framework

- The empirical framework is the meat of the paper

- People are going to read this section closely

- This is where they evaluate your identification strategy

- This is where they evaluate if your paper holds the water of its claims

  - No pressure

- It has two general sections:

1. Estimation strategy
2. Identification strategy

Estimation strategy

- What are you doing to estimate your model?

- Are you running a regression?

  - Difference in differences?
  - Instrumental variables?
  - Twoway fixed effects?
  - Regression discontinuity?
  - Controlling for observables or matching on covariates?

- This is where you stick that all imporant regression equation:

$$y_{it} = \alpha + \gamma D_{it} + X_{it}\beta + \epsilon_{it}$$

where y_{it} is the outcome of interest for individual i at time t, D_{it} is the treatment indicator, X_{it} is a matrix of controls, and ϵ_{it} is the error term

- You do not need to add every control variable here -- just the most important variables to the identification strategy

- Then you can add an extra term that includes all the other controls

Inference (standard erorrs)

- It is critical to add a subsection of the hypothesis test and how you'll draw inferences about it

    - $H_0 : \gamma = 0$, $H_A : \gamma \neq 0$
    - $H_0 : \gamma \leq 0$, $H_A : \gamma > 0$
    - $H_0 : \gamma \geq 0$, $H_A : \gamma < 0$

- Once you have the hypothesis test in place, its time to talk standard errors

- Check out Abadie et al. (2017) for a great discussion of standard errors

- Are you clustering? (Often should be)

- Are you using robust standard errors? (Probably should be)

- Did you use sampling weights (Depends)

- **This is less critical to an inspectional reader**

Identification strategy

- Identification's meaning is a little clunky, but we mean *causal* identification

- You're answering the question: "What assumptions turn this partial correlation into a causal effect?"

- If you have experimental variation with balance tests essentially confirming random assignment, then you're in luck! (Keep it snappy)

- If the balance tests are not perfect, then just explain what you'll control for to make the treatment as good as random (Keep it snappy)

- If you do not have experimental variation, then you'll have to do a bit more work

    - Give an intuitive explanation
    - Discuss the three sources of statistical endogeneity (bias): Omitted variables, reverse causality, and measurement error
    - Ideally include a test of stable unit treatment value assumption (SUTVA) (see Bellemare paper)
    - Do not lie to the reader about the identification strategy

- This subsection is critical to an inspectional reader

Results and Discussion

Results and Discussion

- This is often the least read section of the paper!

- That's cause you put the best parts in the introduction and the empirical framework

- Instead, most readers will skim this for table and figure references and move on

Order of Results

- Start with most parsimonious results (fewest controls) and move to most complex

- This lets you show the robustness of your main results to oddities in the data

- Perhaps shows some unique heterogeneity in the data too!

- Really well-identified papers can do tons of treatment effect heterogeneity tests here

- Others will just do subgroup analysis

Tables

- Tables are the most common way to present results

- Here are some quick rules:

  - They should be self-contained (good title, nice caption, etc.)
  - Coefficient estimates and standard errors should be the same decimal places (2-3?)
  - Omit controls if there are too many, but list whether are included or not
  - Yes/No for fixed effects is important too
  - Add a handful of relevant statistics at the bottom of the table: N, R^2, F-statistic, etc.
  - Estimation results for the estimation sample
  - Covariates should be actual words like "Years of education" not "Edu"

- What do you want to see when you read a paper?

Great table

Table : Ozone and Weather in New York City, 1973

	Ozone (ppb)				
	(1)	(2)	(3)	(4)	(5)
Solar Radiation (Langleys)	0.1148	0.0522	0.1078	0.0509	0.0561
	(0.0234)	(0.0202)	(0.0329)	(0.0236)	(0.0232)
Wind Speed (mph)		-3.109		-3.289	-3.048
		(0.7986)		(0.7777)	(0.8216)
Temperature		1.875		2.052	
		(0.3671)		(0.2415)	
R^2	0.31974	0.63686	0.58018	0.81604	0.82896
Observations	111	111	111	111	111
F-test	2.8202	7.5162	8.2920	19.011	13.218
Month fixed effects	✓	✓	✓	✓	✓
Day fixed effects			✓	✓	✓
Temperature × Month					✓

Notes: The data is a daily time series of air quality measurements in New York City from May to September 1973.
Source: Data from the New York State Department of Conservation (1973)

Tools to automate table creation

- In R:

  - `etable()`
  - `stargazer()`
  - `xtable()`
  - `kable()`
  - `huxtable()`

- In Stata:

  - `esttab`
  - `outreg2`
  - `estout`

- In Python:

  - `pandas` has a `to_latex` method
  - `stargazer` in `rpy2` (R in Python)

Robustness checks

- Dive into robustness checks only once you've shown the main results

- What if you measured the outcome differently in your model? Do the results hold?

- What if you changed your definition of treatment because of some measurement error? Do the results hold?

- Is there a placebo test you can run: i.e. randomly assign fake treatments and see if you get the same results

Mechanisms

- Sometimes your causal inference strategy could have a couple explanations

- It might be perfectly sound on its own that your treatment D caused y, but what if a policymaker can modify D?

- For example, if D is an increase in number of rich friends, policymakers can't change that.

    - But they can perhaps change the amount of income segregation if that's the mechanism at play

- There is no one size fits all way to identify mechanisms

- It is hard!

Limitations

- Tell them the limitations of your paper

- What can't you answer that you know you'd want to answer

- Is this cause of a data limitation?

- A knowledge limitation?

- A measurement limitation?

- A theoretical limitation?

- A scope limitation?

Conclusion

Conclusion

- Summarize what you told them

- Limitations can be put here

- Implications for policy

- Impliations for future research

Title, Abstract, introduction

Title

- You'll know a good title when you see It

- *Science* and *Proceedings of the National Academy of Sciences* say your title should state the results

- But sometimes that can get a bit long, so skip some things like:

  - "A Semiparametric Investigation of Income: Consumption Dynamics"
  - "Nonparametric Evidence from the PSID: Why families matter for the development of children's abilities"

- If your ask a question, "What is the effect of D pon y?" then you could go with,

  - "The effect of D on y: Evidence from [XYZ] context"
  - "Increases in D cause welfare declines from y"

- But really this is hard to tell you!

- Being funny is dangerous, but can pay off if you nail it

Introduction

- Good introductions follow a formula introduced by Keith Head

1. Hook: What is interesting about ths topic?

2. Question: State the question

3. Antecedents: Mention prior work that is critical to understand contribution of your paper[1]

4. Value-added: Describe approximately 3 contributions of your paper

5. Road-map: Outline the organization of the paper (actively)

[1] We all stand on the shoulders of giants.

Hook

- Grab that inattentive inspectional reader's attention

- Make sure it has wide appeal to many types of readers

- Link it to the real world (if possible):

    - Digital payment systems are increasingly common throughout the world
    - They reduce transaction costs and increase efficiency of transactions and are quickly boxing out paper cash
    - The unbanked will struggle to participate in the digital economy
    - What are the welfare consequences?

- Do not bait-and-switch! If your paper cannot address a topic do not!

Take two minutes and write a hook for your paper

- Share it with the class

Question

- Okay, you've got their attention, now what?

- State your research question as clearly as possible

- "Does the elimination of paper cash payments decrease the economic participation of the unbanked?"

- This could be stated even better, but it is moving in the right direction

Take two minutes write the question that naturally follow from your hook

- Is it your question?

Antecedents

- What is the prior literature that is critical to understand your paper?

- Tell a bit of the story of the literature -- how did it go from its earliest work to your contribution?

- Don't discuss papers that are interesting, but tangential

- Stick to the seminal papers and the papers that are most similar to yours

- Don't insult them, but raise the questions they left unanswered

Take two minutes and list the antecedents for your paper

- Do you know what these papers are? Do you think you're missing any?

Value added

- What are the (likely) three contributions of your paper?

  - New measurement of an old economic phenonmenon?
  - New economic phenomenon?
  - New methodological approach to an old economic phenomenon?
  - A new experiment with stronger internal validity?
  - A better identification strategy?
  - First analysis of a new dataset or policy change?

- Dangerous to miss with this though! It is why it is really important to know your antecedents

Take two minutes and list the value added of your paper

- Share it with a partner

Roadmap

Typically this gets pretty boilerplate:

"The remainder of this paper is organized as follows. Section 2 presents the theoretical framework used to study the research question and derives this paper's core testable prediction. In section 3, the empirical framework is presented, first by discussing the estimation strategy, and then by discussing the identification strategy. Section 4 presents the data and discusses some summary statistics. In section 5, the empirical results are presented and discussed, followed by a battery of robustness checks and a discussion of the limitations of the results. Section 6 concludes with policy recommendations and suggestions for future research."

- You won't have your paper rejected over your roadmap

- Just delete it/rewrite it later if a reviewer/editor/me tells you to

- Basically, this is where you make sure everyone is on the same page for where you are headed next

Abstract

- Hopefully, everybody reads these (spoiler: some skip them)

- Typically, this is 100 words or less

- You can often use the introduction as a template

  1. First sentence of the hook
  2. Research question
  3. Value added sections
  4. Polish the above

- Keep this intelligible to any smart, college-educated person

  - Yes, you may need to leave in some esoteric terminology, but keep it minimal

Take two minutes and write an abstract for your paper

- Read it out loud to the class

- Treat this as an elevator pitch for your paper

Literature Review and Background

Literature Review

- Antecdents in the intro are like a mini-literature review

- Many modern economic articles skip a formal literature review -- they just fold it into the intro

- This is partially cause the literature review is often the least read section of the paper

- And lots of people are horrible at writing stimulating literature reviews (yeah, I said it)

- Still if you gotta write one, here's how to do it (and this is good for literature too)

Literature Review organization

- Organization by:

  - Methodology
  - Results/Question (do they disagree? agree?)

- Never by chronology

- Never say, "Papers generally say X"

- Don't refer to author's first names unless you are using them anecdotally in a story of the paper

  - "In a recent paper, Coombs (2023) finds that..." <-- great
  - "In his 2023 paper, Dr. Kyle Coombs of Bates College finds that..." <-- bad

Background

- You may need to add background to your paper

- Institutional settings (pandemic UI policy, Uber surge pricing, bitcoin adoption, etc.) are varied and complex

- If a reader does not understand them, they may not understand your paper

- You just have to keep it as accessible as possible though

Next classes: One-on-ones and Student proposal presentation