

# Writing Tips for Ph. D. Students

John H. Cochrane<sup>1</sup><sup>2</sup>  
Graduate School of Business  
University of Chicago  
5807 S. Woodlawn  
Chicago IL 60637.  
773 702 3059.

[john.cochrane@gsb.uchicago.edu](mailto:john.cochrane@gsb.uchicago.edu)

<http://gsbwww.uchicago.edu/fac/john.cochrane/research/Papers/>

June 8, 2005

<sup>1</sup>Always put your contact info on the front page so that people can find your paper and send you comments! It's the 21st century – get a web page. If your paper is ready for a faculty member to read it, it should be on your webpage. Put the date on the paper so people know if they are reading a new version.

<sup>2</sup>I thank Toby Moskowitz for helpful comments.

# 1 Organization

Figure out the one central and novel contribution of your paper. Write this down in one

## Introduction

The introduction should start with what you do in this paper, the major contribution. You must explain that contribution so that people can understand it. Don't just state your conclusion: "My results show that the pecking-order theory is rejected." Give the fact behind that result. "In a regression of  $x$  on  $y$ , controlling for  $z$ , the coefficient is  $q$ ."

The first sentence is the hardest. Do not start with philosophy, "Financial economists have long wondered if markets are efficient." Do not start with "The finance literature has long been interested in  $x$ ." Your paper must be interesting on its own, and not just because lots of other people wasted space on the subject. Do not start with a long motivation of how important the issue is to public policy. All of this is known to writers as "clearing your throat." It's a waste of space. Start with your central contribution.

Three pages is a good upper limit for the introduction.

I don't write a "roadmap" paragraph: "Section

## Body of the paper

Your task now is to get to the central result as fast as possible. Most papers do precisely the opposite: They have a long motivation, a long literature review, a big complex model that then gets ignored, descriptive statistics, preliminary results, a side discussion or two and then finally Table 12 of “main estimates.” By then we’re all asleep.

Here’s the rule: There should be nothing before the main result that a reader does not need to know in order to understand the main result.

## Theory

In most papers, the “main result” is empirical. There may be some theory or a model, but if you (or the editor!) ask “does this paper expand our knowledge of economic theory?,” the answer is “no.” The theory is there to help understand the empirical work. Following the rule, then, the theory must be the minimum required for the reader to understand the empirical results.

Do not write a “general” model and then “for the empirical work, we now specialize the general shock process to an AR(1), we use only 2 firms rather than a continuum, we assume agents have quadratic utility,” etc. Work out only the specialized model that you actually take to data.

## Empirical work

Start with the main result. Do not do warmup exercises, extensive data description (especially of well-known datasets), preliminary estimates, replication of others’ work. Do not motivate the specification that worked with all your failures. If any of this is really important, it can come afterwards or in an appendix.

You will mightily resist this advice. If you can’t follow it, at least do not put anything before the main result that a reader does not need to know in order to understand the main result.

Follow the main result with graphs and tables that give intuition, showing how the main result is a robust feature of compelling styl



and compare it to others' procedures" at the micro level as well as the macro level. For example, in describing a data transformation, just start with, say, "I adjust income by the square root of household size". Then tell us why adjusting is important, and then talk about different adjustment functions. Most writers do all this in the reverse order.

Previews and recalls are a good sign of poor organization. "As we will see in Table 6" "Recall from section 2" "this result previews the extra analysis of section 4" all often mean you didn't put things in the right order.

Strive for precision. Read each sentence carefully. Does each sentence say something, and does it mean what it says?

Document your work. A fellow graduate student must be able to sit down with your paper and all alone reproduce every number in it from instructions given in the paper, and any print or web appendices. The usual student paper falls short here. There is a sea of verbiage, but I can't figure out how the central table of results was computed, how standard errors were computed, how a simulation was conducted, etc.

Simple is better. Most students think they have to dress up a paper to look impressive. The exact opposite is true: The less math used, the better. The simpler the estimation technique, the better.

## Footnotes

Don't use footnotes for parenthetical comments. If it's important, put it in the text. If it's not important, delete it. Parenthetical comments in footnotes usually mean you haven't organized your ideas; you haven't figured out where to put this thought in a proper linear sequence. Do you really want the reader to stop and read this? Then it should be in the text. Do you think the average reader should not stop? Then delete the footnote. Obviously, lots of parentheses are just as bad as lots of footnotes.

Use footnotes only for things that the typical reader genuinely can skip, but a few readers might want to have attached to the current point. Long lists of references, simple bits of algebra, or other documentation are good candidates for footnotes.

## Tables

Each table should have a self-contained caption so that a skimming reader can understand the fact presented without having to go searching through the text for things like the definitions of Greek letters. Don't go nuts here; some captions are longer than the paper. In my opinion, you can leave out details of variable construction and similar items. "Book/market ratio" is fine; you don't have to tell me that you got book values in June from Compustat. The goal is to allow a skimming reader to understand the table, not to substitute for the detailed documentation that must be in the paper somewhere.

The caption of a regression table should have the regression equation and the name of



Much bad writing comes down to trying to avoid responsibility for what you're saying. That's why people resort to passive sentences, "it should be noted that", poor organization with literature first and your idea last, and so on. Take a deep breath, and take responsibility for what you're writing.

Present tense is usually best. You can say "Fama and French 1993 find that" even though 1993 was a while ago. The same goes for your own paper; describe what you find in Table 5 not what you will find in Table 5. Most importantly, though, keep the tense consistent. Don't start a paragraph in past tense and finish it in the future.

Use the normal sentence structure: subject, verb, object. Not: "The insurance mechanisms that agents utilize to smooth consumption in the face of transitory earnings fluctuations are diverse" Instead: "People use a variety of insurance mechanisms to smooth consumption.." (I also changed the starchy "agents" to the concrete "people," and the simple "variety" rather than the fancy "diverse." Actually, this whole sentence probably should be dumped; it was introducing a paragraph that described the mechanisms. It's a throat-

that “this” could point to.

Hyphens are widely misused. Here’s the rule from the JFE style sheet: “Hyphens are used for true compound modi

without saying. I don't list every single place I've given the workshop in the thanks. I'm not ungrateful, but the long list can get out of hand.

Don't start your paper with a cute quotation.

Don't overuse italics. (I use them far too much.) It's best to use them only when the emphasis in a sentence would otherwise not be clear – but maybe then you should rewrite the sentence so that the emphasis really is clear. (Who is that shouting in here?)

When describing the sign of a casual link, one direction is enough. "When Jane goes up (down) on the teeter-totter, Billy goes down (up) on the other side," the stuff in the parentheses is distracting. Add "and vice versa" if you must.

Every sentence should have a subject, verb and object. No sentences like "No sentences like this."

### 3 Tips for empirical work

These tips verge on "how to do empirical work" rather than just "how to write empirical work," but in the larger picture "doing" and "writing" are not that different.

What are the three most important things for empirical work? Identification, Identification, Identification. Describe your identification strategy clearly. (Understand what it is, first!) Much empirical work boils down to a claim that "A causes B," usually documented by some sort of regression. Explain how the causal effect you think you see in the data is identified.

1. Describe what economic mechanism caused the dispersion in your right hand variables. No, God does not hand us true natural experiments very often.
2. Describe what economic mechanism constitutes the error term. What things other than your right hand variable cause variation in the left hand variable?
3. Hence, explain why you think the error term is uncorrelated with the right hand variables in economic terms. There is no way to talk about this crucial assumption unless you have done items 1 and 2!
4. Explain the economics of why your instruments are correlated with the right hand variable and not with the error term.
5. Do you understand the difference between an instrument and a control? In regressing  $y$  on  $x$ , when should  $z$  be used as an additional variable on the right hand side and when should it be an instrument for  $x$ ?
6. Describe the source of variation in the data that drives your estimates, for every single number you present. For example, the underlying facts will be quite different as you

add fixed effects. With firm fixed effects, the regression coefficient is driven by how the variation over time within each firm. Without firm fixed effects, the coefficient is

Of course, every important number should include a standard error.

## 4 Seminar presentations

You will not believe how fast the time will go by.

