Writing Tips for Ph. D. Students

John H. Cochrane[1,2]
Graduate School of Business
University of Chicago
5807 S. Woodlawn
Chicago IL 60637.
773 702 3059.
john.cochrane@gsb.uchicago.edu
http://gsbwww.uchicago.edu/fac/john.cochrane/research/Papers/

June 8, 2005

---

[1]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.

1 Organization

Figure out the *one* central and novel contribution of your paper. Write this down in one paragraph. As with all your writing, this must be concrete. Don't write "I analyzed data on executive compensation and found many interesting results." Explain what the central results *are*. For example, Fama and French 1992 start their abstract with: "Two easily measured variables, size and book-to-market equity, combine to capture the cross-sectional variation in average stock returns associated with market β, size, leverage, book-to-market equity, and earnings-price ratios."

Distilling your one central contribution will take some thought. It will cause some pain, because you will start to realize how much you're going to have to throw out. Once you do it, though, you're in a much better position to focus the paper on that one contribution, and help readers to get it quickly.

Your readers are busy and impatient. No reader will ever read the whole thing from start to finish. Readers skim. You have to make it easy for them to skim. Most readers want to know your basic result. Only a few care how it is different from others. Only a few care if it holds up with different variable definitions, different instrument sets, etc.

Organize the paper in "triangular" or "newspaper" style, not in "joke" or "novel" style. Notice how newspapers start with the most important part, then fill in background later for the readers who kept going and want more details. A good joke or a mystery novel has a long windup to the final punchline. Don't write papers like that – put the punchline right up front and then slowly explain the joke. Readers don't stick around to find the punchline in Table 12.

The vast majority of Ph.D. student papers and workshop presentations (not all by students!) get this exactly wrong, and we never really find out what the contribution of the paper is until the last page, the last table, and the last 5 minutes of the seminar.

A good paper is not a travelogue of your search process. We don't care how you came to figure out the right answer. We don't care about the hundreds of things you tried that did not work. Save it for your memoirs.

Abstract

Most journals allow 100-150 words. Obey this limit now. The main function of the abstract is to communicate the one central and novel contribution, which you just figured out. You should not mention other literature in the abstract. Like everything else, the abstract must be concrete. Say what you find, not what you look for. Here too, don't write "data are analyzed, theorems are proved, discussion is made.."

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 2 sets out the model, section 3 discusses identification, section 4 gives the main results, section 5 checks for robustness, section 6 concludes." It seems a waste of space; readers will figure it out when they get there and I save a paragraph against the editor's page count. Make your own mind up about this question; but realize it's not mandatory.

Literature review

Do not start your introduction with a page and a half of other literature. First, your readers are most interested in just figuring out what you do. They can't start wondering if it's better than what others have done until they understand what you do. Second, most readers do not know the literature. It's going to be hard enough to explain your paper in simple terms; good luck explaining everyone else's too.

After you've explained your contribution, then you can write a brief literature review. Make it a separate section of otherwise set it off so people can skip it who aren't interested. Remember, it will be very hard for people to understand how your paper is different from others' given that they don't understand your paper yet, and most of them have not read the other papers.

Be generous in your citations. You do not have to say that everyone else did it all wrong for your approach and improvements to be interesting.

Literature reviews have gotten way out of hand. It is not necessary to cite every single paper in the literature or to write a *Journal of Economic Literature* style review. The main point of the literature review should be to set your paper off against the 2 or 3 closest current papers, and to give proper credit to people who deserve priority for things that might otherwise seem new in your paper. Some people worry a lot about strategic citations; choosing citations to hint to editors who they should assign as referees and adding loads of citations to make sure referees see themselves. Whatever one thinks of these practices, we can agree you should get rid of all the fluff in the final version.

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 stylized facts in the data. Follow that with limited responses to potential criticisms and robustness checks. Most of those should end up in your web appendix.

Conclusions

Really, a conclusions section should not be necessary. If you did a good job of explaining your contribution in understandable prose in the introduction, and then documenting those claims in the body of the paper, (writing in good triangular style), then saying it all over again is pointless. I tried omitting the conclusions section a few times, though, and this was too radical for editors and referees. It is true that some people skip to the conclusion to

look for the main result, but that's because they are used to authors who don't explain it well enough in the introduction.

Thus, conclusions should be short and sweet. Do not restate all of your findings. One statement in the abstract, one in the introduction and once more in the body of the text should be enough! You can include a short paragraph or two acknowledging limitations, suggesting implications beyond those in the paper. Keep it short though – don't write your grant application here outlining all of your plans for future research. And don't speculate; the reader wants to know your facts not your opinions.

Appendices

Appendices are a great tool. Take that delicious section that has so many insightful comments on the literature, the general version of the model, the 57 robustness exercises that you did, and dump them in to an appendix. This is a good way to get them out of the paper. Eventually you'll dump them out of the appendix too.

Seriously, careful authors, referees and critics often want to document that the main result is robust to various other ways of doing things. You have to do that, but once you've verified that it does not make that much difference and you've found the one best way of doing things in your main result, it isn't worth space in the paper to present all the checks and variations. Appendices are a great way to solve this problem, and you can just summarize all the things you did in the paper. You can put the appendix on your and the journal's website. ("Bond risk premia" with Monika Piazzesi is an example of a web-appendix gone wild.)

2   Writing

Keep it short

Keep the paper as short as possible. Every word must count. As you edit the paper ask yourself constantly, "can I make the same point in less space?" and "Do I really have to say this?" Final papers should be no more than 40 pages. Drafts should be shorter. (Do as I say, not as I do!) Shorter is better.

Don't repeat things. In other words, if you've said it once, you don't have to say it again. Most of all, it uses up extra space and reader's patience to have to see the same point made over and over again. So, once again, repetition is really a bad idea. (Get the picture?!) "In other words" is a sign of trouble. Go back and say it once, right.

General points

Follow the rule "first describe what you do, then explain it, compare it to alternatives,

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

the variables, especially the left hand variable.

No number should appear in a table that is not discussed in the text. You don't have to mention each number separately; "Row 1 of Table 3 shows a u-shaped pattern" is ok. "Table 5 shows summary statistics" (period) is not ok. If it's not worth writing about in the text, it's not worth putting in the table.

Use the correct number of significant digits, not whatever the program spits out. 4.56783 with a standard error of 0.6789 should be 4.6 with a standard error of 0.7. Two to three significant digits are plenty for almost all economics and finance applications.

Use sensible units. Percentages are good. If you can report a number as 2.3 rather than 0.0000023, that's usually easier to understand.

Figures

Good figures really make a paper come alive, and they communicate patterns in the data much better than big tables of numbers. Bad or poorly chosen figures waste a lot of space. Again, give a self-contained caption, including a verbal definition of each symbol on the graphs. Label the axes. Use sensible units. Don't use dotted line types that are invisible when reproduced. Don't use dashes for very volatile series.

Writing tips

The most important thing in writing is to keep track of what your reader knows and doesn't know. Most Ph.D. students assume far too much. No, we do not have the details of every paper ever written in our heads. Keep in mind what you have explained and what you have not.

The reader usually wants most of all to understand your basic point, and won't start criticizing it before he or she understands it. That's behind my advice to first state and explain what you do, and save defending it and comparing it to other approaches until much later.

Use active tense. Not: "it is assumed that $\tau = 3$", "data were constructed as follows.." Gee, I wonder who did that assuming and constructing? Search for "is" and "are" in the document to root out every single passive sentence.

"I" is fine. Don't use the royal "we" on a sole-authored paper. "I assume that $\tau = 3$." "I construct the data as follows." If it seems like too much "I," you can often avoid the article altogether. For example, I think it's ok to write "Table 5 presents estimates" rather than "I present estimates in Table 5", though a purist might object to making a Table the subject of a sentence. I use "we" to mean "you (the reader) and I," and "you" for the reader. "We can see the u-shaped coefficients in Table 5" or "You can see the u-shaped coefficients" is much better than "The u-shaped coefficients can be seen" (passive) or "one can see the u-shaped coefficients" (who, exactly?)

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-clearing sentence that violates the rule that every sentence should mean something. The fact that people use a variety of mechanisms is not big news, the news is what the mechanisms are.)

Avoid technical jargon wherever possible.

Writing should be concrete, not abstract. (Insert concrete examples.)

Little writing tips

Don't use adjectives to describe your work: "striking results" "very significant" coefficients, etc. If the work merits adjectives, the world will give them to you.

If you must use adjectives, don't use double adjectives. Results are certainly not "very novel."

Use simple short words not big fancy words. "Use" not "utilize." "several" not "diverse".

It is usually the case that most good writers find that everything before the "that" should be deleted from a sentence. Read that sentence again starting at "Everything": it's true, isn't it? "It should be noted that" is particularly obnoxious. Just say what you want to say. "It is easy to show that" means that it isn't. Search for "that" in the document to get rid of these. Similarly, strike "A comment is in order at this point." Just make the comment. These phrases also violate the rule that each sentence should mean what it says. Is the point of the sentence really that "it should be noted?" Or is this just a wimpy way to bring up the topic?

Clothe the naked "this." "This shows that markets really are irrational..." This what? "This" should always have something following it. "This *regression* shows that...." is fine. More generally, this helps (no, that should be "this *rule* helps," right?) you to avoid an unclear antecedent to the "this." Often there are three or more things in recent memory

that "this" could point to.

Hyphens are widely misused. Here's the rule from the JFE style sheet: "Hyphens are used for true compound modifiers before the noun (e.g., after-tax income, risk-free rate, two-day return, three-digit SIC code, value-weighted index) unless part of the compound modifier is an adverb ending in 'ly' (e.g., previously acquired subsidiary, equally weighted index, publicly traded stock). When there is no risk of misinterpretation, the hyphen can be omitted, but the treatment must be consistent throughout the paper." Note the hyphen is optional, so you don't have to construct monstrosities like "continually-rebalanced-equally-weighted portfolio." Don't use hyphens in other circumstances, e.g. "The paper focuses on small-stocks."

People forget Greek letter definitions. If you define them once in an obscure part of the text and then use naked references ("$\theta = 3$ gives the best fit") no one will know what you're talking about. Define them clearly in an easy-to-find place. It's best to give them a name too, and then remind people of the name and the number ("I find the best fit when the elasticity of substitution θ equals 3.") This is the one place where a little repetition isn't bad. If you've reminded them of the name in the last paragraph or two, however, you can use the naked letter.

Strike "I leave x for future research." We're less interested in your plans and excuses than we are in your memoirs.

Never use the words "illustrative test" or "illustrative empirical work." Never do illustrative work. Do real empirical work or don't do any at all. Illustrating technique with empirical work you don't believe in is a waste of space. Even if you do it, there is no faster way to get readers to fall asleep than to tell them that what you're doing doesn't really matter.

You don't need to "assume" things about a model. Don't write "I assume that consumers have power utility" (And, of course, don't write "it is assumed that utility is power," right?) You are describing a model, not reality, so you can just state the model structure. "Consumers have power utility." ("In this model" is understood.) Save "assumptions" for things that really do modify the real world, "I assume there are no shifts in the demand curve so that the regression of price on quantity identifies the supply curve."

Keep down the number of clauses in your sentences, and the number of things kept hanging.

"Where" refers to a place. "In which" refers to a model. Don't write "models where consumers have uninsured shocks," write "models in which consumers have uninsured shocks."

Don't abbreviate authors' names, "FF show that size really does matter." There is always enough space to spell out people's names. You'd want them to write out yours, no?

It is appropriate to thank people who have helped you in the author footnote. I don't add the qualifier about not blaming people I thank for comments for mistakes. It goes

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 (mostly) driven by variation *across firms* at a moment in time.

7. Are you sure you're looking at a demand curve, not a supply curve? As one way to clarify this question, ask "whose behavior are you modeling?"

    Example: Suppose you are interested in how interest rates affect housing demand, so you run the number of new loans on interest rates. But maybe when housing demand is large for other reasons, demand for mortgages (and other borrowing demand correlated with demand for mortgages) drives interest rates up. You implicitly assumed stable demand, so that an increase in price would lower quantity. But maybe the data are generated by a stable supply, so that increased demand raises the price, or some of both. Are you modeling the behavior of house purchasers or the behavior of savers (how savings responds to interest rates)?

8. Are you sure causality doesn't run from y to x, or from z to y and x simultaneously? Think of the obvious reverse-causality stories.

    Example: You can also think about the last example as causality: Do interest rates cause changes in housing demand or vice versa (or does the overall state of the economy cause both to change)?

9. Consider carefully what controls should and should not be in the regression. Most papers have far too many right hand variables. You do not want to include all the "determinants" of y on the right hand side.

    (a) High R^2 is usually bad – it means you ran left shoes = $\alpha + \beta$ right shoes $+\gamma$price + error. Right shoes should not be a control!

    (b) Don't run a regression like wage = a + b education + c industry + error. Of course, adding industry helps raise the R^2, and industry is an important other determinant of wage (it was in the error term if you did #2). But the whole point of getting an education is to help people move to better industries, not to move from assistant burger-flipper to chief burger-flipper.

Give the stylized facts in the data that drive your result, not just estimates and p values. For a good example, look at Fama and French's 1996 "Multifactor explanations." In the old style we would need one number: the GRS test. Fama and French show us the expected returns of each portfolio, they show us the beta of each portfolio, and they convince us that the pattern of expected returns matches the pattern of betas. This is the most successful factor model of the last 15 years ...even though the GRS test is a disaster! They were successful because they showed us the stylized facts in the data.

Explain the economic significance of your results. Explain the economic magnitude of the central numbers, not just their statistical significance. Especially in large panel data sets even the tiniest of effects is "statistically significant." (And when people show up with the usual 2.10 t statistic in large panel data sets, the effect is truly tiny!)

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

4 Seminar presentations

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

Since time is limited, it's especially important to get to the point. We can't skim to the important stuff in a seminar!

You don't need any literature review or motivation in a seminar. Just get to the point. Gene Fama usually starts his seminars with "Look at table 1." That's a good model to emulate.

Don't "preview" results. It wastes time; why say it twice rather than say it once, right?

Don't make slides with a bullet point for every word you intend to say. This forces you into a preplanned order, and then you can't change on the fly when you figure out how fast time is going by. Slides are fine that only contain equations, tables and graphs – things we really need to *see*. At most use words for the one or two really important things you want people to know, e.g. "Identification: interest rates do not respond to fiscal shocks in the Ricardian model." Also, you want people to remember the structure of the model, definitions of variables, etc. If you have too much junk on the slides, people can't see the utility function while you're talking about the production function, so they get lost. People don't remember equations from one slide to the next.

You have to leave slides up for a decent amount of time in order for people to digest them. That means you will not be able to put up 1 slide per minute!

As in writing the paper, your main objective is to get to the #1 important contribution as fast as possible.

Most seminars are a disaster. They start with pointless motivation and policy implications, which the audience can't follow since we don't know the result. Then we get a long literature review, which is even more boring since we don't know the point of this paper much less what everyone else did. Then we get a results preview. Usually, the presenter says "I'll preview the results now because I may not have time to get to them all," a strangely self-fulfilling prophecy. Since showing the main results is the only reason you came, why not just start *right now!* Worse, the reason we run out of time is because we wasted half an hour on the stupid preview! The seminar then bogs down as people start asking questions about the previewed results; most of the questions are dumb ("I measure the demand elasticity at 0.3." "But how did you identify supply shifts?") since they will be explained in a proper presentation of the results. But the questions are totally reasonable since the claim with no documentation is meaningless. Next, we get (in empirical papers) some "theory" that is really beside the point and only serves to provoke more needless argument (no, there really is no way to distinguish the "behavioral" and "rational" explanation. Clever audience

members will come up with stories that reverse all the signs.) Then we get some distracting preliminary results and tables and graphs of unrelated observations. More pointless discussion erupts; people don't know what point the speaker is trying to make and the discussion goes off in to tangents. Finally the speaker sees there is only 10 minutes to go, tells people to be quiet, and the main results go by in a big rush. Everyone is tired and confused and doesn't follow anything. I timed the finance workshop last winter quarter and *not one* paper got to the main results in under an hour!

Listen to the questions, all the way to the end, then count to three before answering. Yes, you're in a rush, and yes, you think you can guess what the question will be and you know the answer. This isn't a game show, and much of the time you actually don't know what the question will be.

Keep a sheet of paper handy. You may not have a quick answer to every question, and some questions may point to good things to change in the paper.

You cannot make it too simple. Most presenters, especially Ph. D. students overestimate dramatically how much theory people can digest in one sitting, and how quickly they can memorize and digest models and results.

Speak loudly, slowly and clearly.

There's nothing wrong with ending early!

5 Conclusion

May economists falsely think of themselves as scientists who just "write up" research. We are not; we are primarily writers. Economics and finance papers are essays. Most good economists spend at least 50% of the time they put into any project on writing. For me, it's more like 80%.

Pay attention to the writing in papers you read, and notice the style adopted by authors you admire.

I got a lot out of reading William Zinsser's *On Writing Well,* and D. McCloskey's *Rhetoric of Economics.* I also found Glenn Ellison's "The slowdown of the economics publishing process" in the JPE useful for thinking about how papers should be structured (and refereed and edited, but that's another story).