Ilyana Kuziemko
Princeton Economics
Writing tips for Ph.D. students
Last Updated: December 2023

Learning to write more clearly and succinctly is a lifelong process. I am often horrified when I read my past writing, which I like to think means that my writing continues to improve.

This document includes a list of tips I have found useful, as well as a list of the most common mistakes I see in academic writing.

Preliminaries

Use LaTeX for writing papers. You may incur some upfront costs of learning its various tricks, but they will pay out over the rest of your career. Plus, while unfair, some readers, editors and referees look down on papers that are written in Word or some other non-LaTeX platform.

For the same reasons, learn LaTeX beamer for presentations.

For creating summary statistics and regression tables, I would strongly suggest using Stata packages such as esttab (or the equivalent in R) that "play nicely" with LaTeX. When you compile the document in LaTeX, your tables and figures should automatically update.

For figures you draw (e.g., to illustrate a result from a model) as opposed to generate via data, do not use PowerPoint. LaTeX offers a package called tikz that produces *beautiful* figures.

I have heard that ChatGPT is very helpful in writing some basic esttab and tikz code, though I haven't tried it myself!

Some rules related to style, usage and grammar

1. *This* cannot be used as a subject in academic writing. "This shows that…" is *verboten*. "Taking this into account, I will not consider children under five" should be replaced with something like "Taking into account the low response rates of children under age five, I will not consider them." Using *this* as a noun allows you to slip into vague and sloppy writing. Formal writing requires that you "clothe the naked *this*," even though doing so will often result in longer sentences.

2. Anytime you use words like *greater, less, more, better, worse*, make sure you spell out the comparison. "This dataset will allow me to more accurately determine household composition…." More accurately than *what*?

3. Avoid assertions. Even in your introduction, you must provide citations for almost any claim you make.

4. Avoid advocacy. Even for more informal research proposals, I would avoid language such as "I will hope to show…."

5. You should not use contractions (e.g., "can't," "won't," "didn't") in formal writing.

6. Economists use the present tense whenever possible. For example, even though you likely cleaned your data before you ran regressions, you can use the present tense when you describe both processes.

7. Even for more informal writing such as research proposals, type out *and* instead of using & and or instead *of* /. Write "200 or more" and not "200+." After all, you are writing a paper, not an email. ☺

8. Punctuation marks appear *inside* of quotation marks in standard American usage (British usage has different rules, so you will see punctuation outside of quotation marks in *The Economist* and other British publications).

9. In standard American usage, it is correct to write US (no periods) or U.S. (with periods) when you are using the term as *a proper noun*, but one *must* write U.S. (with periods) if you are using the word as an adjective. So, "crime fell in the 1990s throughout the US" and "crime fells in the 1990s throughout the U.S." are both correct.

But "U.S. policy toward Iraq" is correct while "US policy toward Iraq" is incorrect. Again, standard British usage rules are different on this point.

10. In standard American English, *you may not separate two independent clauses with a comma*. This violation even has its own name ("comma splice")! Other languages allow comma splices in formal writing (you see it a lot in French), but not in American English. "I first normalized each variable, I then re-estimated the regression" is incorrect and can be fixed by simply dividing into two sentences or adding "and" in front of "I then re-estimated."

11. Learn the difference between *that* versus *which*. "I standardized the variables that were provided by the Department of Education" implies that there may be *other* variables that you did *not* standardize. "I standardized the variables, which were provided by the Department of Education" implies that you standardized all the variables, and, oh, by the way, the variables came from the DOE. In the first example, "were provided by the Department of Education" is absolutely required to make sense of the sentence, whereas in the second it could be cut without the reader being misled.

12. When capitalized, the words Republican, Democrat, Liberal and Conservative refer to political parties (for example, the first two are political parties in the US, the second two are political parties in the UK). When *not* capitalized, these words are meant in a more literal sense (e.g., "republican form of government," "democratic principles," "liberal trade policies," "conservative traditions").

13. The Latin phrase *et alia* means "and others" and is abbreviated *et al.*, with a period following *only* the second word.

14. "Small" numbers should be written out ("the first nine variables," not "the first 9 variables"). Different style guides disagree on what constitutes "small," but there is strict agreement that any number below ten is "small." Another version of the rule is that any number that does *not* require a hyphen should be written out. So, 23 (not "twenty-three"), but "fifteen" (not 15) and "fifty" (not 50). But, in any case, certainly do not use digits for numbers below ten.

15. Here is another picky rule regarding numbers. It is bad form to start a sentence with a number. "14% of subjects dropped out before the final round" should be re-worked, even if it is a bit forced. "By the final round, 14% of subjects had dropped out" would work.

16. One last picky rule regarding numbers. A decimal less than one should always include the leading zero. "These exclusions lead us to drop .42 percent of subjects" should be "drop 0.42 percent of subjects."

17. If you are describing something that is *not countable*, use "less" or "amount of" or "level of." For example, "I have noticed less cynicism among my students than in past years." Similarly, "I always underestimate the amount of time required to grade papers." However, you *cannot use these modifiers if you are describing something that is, in principle, countable* (even if in practice it would be hard to actually count). For example, "surveys show that fewer Americans are uninsured in 2016 than in 2010" is correct. But "less Americans are uninsured" is *incorrect (and sounds awful to boot)!*

18. *Data* are plural: so, "this data is" is incorrect, whereas "these data are" is correct. *Dataset* is singular.

19. *Its* is the possessive form of *it* and does *not* have an apostrophe. *It's* is the contraction for "it is."

20. *Cannot* is one word. "I cannot link parents to children in this dataset," is correct. "I can not link" is incorrect.

21. The use of *between* implies *exactly two* items. When describing more than two items, you must use *among*. "Electoral votes are distributed between states based roughly on their population" is incorrect. "Electoral votes are distributed *among* states" is correct.

22. Avoid using a noun as an adjective (e.g., "grammar error" is incorrect, while "grammatical error" is correct). Nouns are often used as parts of compound adjectives, in which case they must be hyphenated, but *only when they precede the noun* (so "race-based policy," but "the policy is race based").

Formatting

Unless asked otherwise, you should use 12-point Times New Roman font, one-half spacing, and one-inch margins. Do not assume your reader has the same young, sharp eyes that you do!

Please assume your reader will print your paper on a black-and-white printer and thus do *not* rely on color to make any distinction in either the text or in figures. Please also follow this rule for presentations (eight percent of men are colorblind).

Presenting tables and figures

Make tables and figures as "self-contained" as possible. For your actual job paper, the first people who will read it are the members of the junior recruiting committee, who will each have about 100 papers to read. They will likely read the abstract, intro, some of the text, and then flip to tables and figures.

You want to *obsess* about these tables and figures. On the one hand, you want someone to understand the table or figure without referring back to the text. On the other hand, you cannot overwhelm them with information in that table or figure. Getting the balance right takes a lot of thought. Descriptive (but judiciously curated) notes under the table or figure are essential.

One trick is to imagine that someone is teaching your paper to graduate students. Could that person put the figure in their slides and have the grad students follow along without too much confusion?

A few tips I have found useful in creating tables are below.

1. The dependent variable should always be made clear, either as a column heading if the dependent variable changes across specifications, or just as a prominent heading somewhere toward the top of table if it is common for all specifications (the Stata esttab option "mgroups" is helpful for creating multi-column headings, as are the LaTeX commands \multicolumn and \cmidrule).
2. For regression tables, I almost always put the mean of the dependent variable for each column in one of the rows in the table footer. In some cases, the standard deviation might also be useful. If the sample and dependent variables are common across all specifications, then I would include this information somewhere prominent in the table notes. Including these summary statistics will help readers judge the effect size of your coefficients of interest without having to refer back to the text.
3. Do not ever put variable names in your table! Put reader-friendlier variable *labels* instead. So, use "Logged family income," not "log_fam_inc." In esttab, simply use the "label" command. Having variable names instead of labels looks lazy and sloppy and is obviously much more taxing for the reader.
4. The package *booktabs* makes tables more attractive in LaTeX.