The Recipe for a Successful Thesis in Applied Economics[*]

Joakim Westerlund[†]
University of Gothenburg
Sweden

August 23, 2010

Abstract

This paper provides tips of things to think of when writing a thesis in applied economics. One of the most important ingredients in the recipe for success is to get an early start, and to begin looking for data as soon as possible.

1 Introduction

What is "research"? The traditional view is that research is "the search for knowledge," which reflects the idea that science should be completely objective. The researcher simply collects the data, which then "speak for themselves." Figurly speaking, knowledge is fruit on a tree in the forrest, and all the researcher needs to do is to discover the tree and collect the knowledge (Greenlaw, 2006).

Of course, this is not entirely true. What is missing here is a distinction between knowledge and facts. Knowledge is our common understanding of how things work, and if we do not know exactly, it is our best guess. Facts, on the other hand, are just data. In other words, knowledge is facts with meaning; it is the researcher's best interpretation of the facts, and research is the creation of such knowledge.

But different researchers interpret the same facts differently, and so there is not just "one" knowledge. Rather, knowledge is a dialog between researchers as they strive to strengthen their interpretations of the facts. Through this dialog weaker interpretations, or arguments, are gradually abandoned, while the stronger ones are refined and made even better.[1] Weaker

---

[†]Telephone: +46 31 786 5251, E-mail address: `joakim.westerlund@economics.gu.se`.
[1]An argument is a point of view on a question.

arguments are those with less supporting evidence. Since the data are also the facts, having weak support means that the argument does not fit the facts and that it must be flawed in one way or another.

Many students believe that they do not know enough to be able to conduct their own research. One reason for this is that they believe that research is only made up of ground-breaking discoveries like the theory of relativity or the cure for cancer. This is wrong, wrong, wrong! Indeed, the majority of research makes only very marginal contributions to our current understanding of things, and it is almost never made directly from scratch, but is made based on the work of others. So contributions are marginal, and you do not even have to work everything out yourself – easy! There is therefore no reason for why students should not be able to do research.

2   Getting started

According to Greenlaw (2006) the following items are crucial when starting up a new thesis project:

1. A good question.

2. A testable hypothesis.

3. A good data set.

4. An econometric approach that fits the problem and that can be used to test the hypothesis of interest.

It is important to begin thinking about these items right away, before you begin the actual thesis work, so that once you begin you know what to do. Do not just start to think about the question. Research is a highly nonlinear process that must be carried out in an iterative, rather than a step-wise, fashion. For example, although you need to begin to think about the question early (step 1), you also need to think about the availability of data (step 3). If there are no data, then you can just as well start looking for something else. In fact, as soon as you have just a partial idea of what to do you should begin to look for data. Similarly, in formulating the hypothesis (step 2), you need to know what kind of data you have (step 3).

2.1 The question

Prize-winning research questions are difficult to come by. In fact, most researchers have to settle with much less. The most important thing is that the question is interesting to you. It should of course also be interesting to your audience, but if it is interesting to you then you are also highly motivated and engaged, which tends to show in the way you write. Your audience in this case are those that perform research on the topic, which also includes students.

The best way of getting started is therefore to pick a topic that interest you, ideally one that you have some background in. You might for example have taken a course on a subject that you liked. Once you have homed in on a topic you start looking for articles. The authors of these articles are your audience; they are experts and therefore they know what is new and "hot" in the area. These guys know the econometric tools that fit the problem best, and most of the time they also provide at least some overview of the literature. So if you found the article interesting and want to know more, then you know where to look. You also get the data sources, which can be a valuable piece of information, especially if you are new in the area.

Reading articles is therefore a very good way to start, if not the best! By surveying the literature you get a good picture of where the research frontier is, what is known, and what questions that have not been explained in a satisfactory way. In most cases you will find that there are large differences in how researchers choose to interpret the facts, and there is almost always conflicting findings. Researches have different backgrounds, use different data sets and different econometric tools, and therefore the conclusions they reach will also be different. This is good because it leaves the door open for more research, which is where you might decide to step in.

But there is also the possibility that the arguments of the article are flawed in one way or another. In fact, this is more of a rule than an exception. Here are some questions to help you evaluate an author's argument, and to identify potential shortcomings:

1. Does the question make sense? It the motivation convincing?

2. Are the data used appropriate for the task?

3. Is the econometric approach satisfactory? Does it test what it is supposed to?

4. Are the assumptions reasonable? This applies to the both the formulation of the question, which is usually rooted in economic theory, and the econometric analysis.

5. Does the conclusions follow from the evidence presented, or is the author overstating the results?

Another way to start is to try to mimic the study of an article that you like. It might for example be interesting to use the same econometric approach and to pose the same question but for another set of countries, or for another time period. This is a "safe" way to plan your thesis because it means that you know that the question is interesting, and that the econometric approach is suitable. You also get a good feeling for how to structure and write up the thesis. Another approach is keep the data but you replace the econometric approach with one that you think is better suited. This is also quite "safe" in the sense that you can contact the author directly and ask for the data, which is then already assembled and ready to go.

As for the search for interesting articles, although the library might seem as the most natural place to start, most of the time it is actually better to simply stay at home and search the Internet. Google (http://google.com) is an excellent research companion that should not be forgotten when starting up a project. The best way is to combine keywords. You might for example combine a keyword from the economic topic with one from the econometric approach that you would like to work with. The problem is that not all hits will be scholarly in nature, and it might be difficult to separate the corns from the weeds. One thing to keep in mind here is that while most people post their material as text on the Internet as separate webpages, researchers usually upload their work as Adobe pdf or Word documents, which are easily identified in Google.

Another thing to keep in mind is that most, if not all, researchers have their own personal webpages, where they post – among other things – working papers and data. Hence, once you know just a little about the literature and the researchers that write there, you can start to search directly for their names and affiliations, which will speed up the search.

Of course, not all articles are easy, and there are several obstacles that you might come across. The terminology or jargon might be unfamiliar to you, and the mathematics might be difficult. There is also the econometrics, which is likely to be at least a little novel to you. But do not let this stand in your way! The jargon will oftentimes become clearer as you

continue reading, thereby putting everything in context. The math is only for giving rigor to an argument that can be explained in words. Think of it as a "proof" of how the conclusion was reached. But the proof is not necessary if you only want to understand what is being proven. If you feel uncomfortable with the econometrics, then it is just to look it up in one of your old econometrics text books.

2.2   The hypothesis

The research question should boil down to a testable hypothesis that can be brought to the data and tested. This is crucial. At the same time, it is usually nothing that you have to be terribly worried about. The reason is that once you start reading articles you will realize that most questions have already been hypothesized.

2.3   The data

A good data set is key, and is oftentimes what makes the difference between a good and a mediocre thesis.

As soon as you have an idea of what to do you should start searching for data. One cannot stress this point enough. Many students are very enthusiastic about their topic and through themselves into the writing process, spending weeks on getting the arguments on paper, only to find out that there are no data available. This is not only time-consuming but also devastating for morale, because being forced to start all over again is never funny. Do not make the same mistake! The following quotation, taken from Greenlaw (2006, page 139), illustrates the point:

> "It is a capital mistake to theorize before one has data. Insensibly one begins
> to twist facts to suit theories, instead of theories to suit fact."

Sir Arthus Conan Doyle

Thus, it is never too early to begin looking for data. A good tip is to pay attention to the data used in the articles you read. What is the structure of the data? Are they cross-section, time series or maybe panel data? If they are time series data, then what is the frequency – annual, quarterly, monthly, or daily? What kind of variables do you need? Some articles use a lot of variables, most of which are there to control for outside influences that might otherwise affect the test of the hypothesis of interest. Such controls are usually not crucial

and can in many cases be easily replaced. The immediate task is therefore to identify those variables that are necessary for the test, and then to check the availability of the data.

Some variables are very easy to find but others are harder to come by. For example, while aggregate macroeconomic and financial data are very easy to find, detailed data of household behavior are more rare.

Many students are able to find most of the data they need, but some variables are missing. In such cases you might want to consider replacing the missing variables by suitable proxies. A proxy is a variable that might be defined differently but that behaves very much like the variable that you really want. Suppose for example that the article that you have been reading uses gross domestic product (GDP), but the data you find are incomplete, and have a lot of missing observations. Then you might try to use either gross national product or industrial production as a proxy. These have the same trend and major turning points as GDP, and are therefore expected to produce roughly the same test results.

In the unlikely event that you have read all the articles you can find but still do not know where to look for data, then Resources for Economists on the Internet (http://rfe.org) is a good place to start.

Once you have found the variables you need, the next step is to check how many observations you have. Many variables are not measured in the same way, and might not even cover the same units and/or time periods. What is considered as a decent sample size varies across topics or disciplines, but usually you need at least 30 observations or so to be able to test your hypothesis without running into degree of freedom problems.

2.4 The econometric approach

Once you have secured some data, which do not necessarily have to be the complete set, the next order of business is to obtain some preliminary results. This should be done as quickly as possible. The purpose is to get a first glance of the evidence. Do the results look reasonable, and how do they conform to your hypothesis?

Of course, this is not to say that you should only proceed in case the results look favorable, or that you should alter your hypothesis so that it is confirmed by the testing. It only means that since the data are there it seems like a waste not to make use of the results from the beginning.

Knowing something about the outcome of the empirical analysis can be extremely valu-

able for at least three reasons. Firstly, as an insurance, it means that you can continue working without having to lie awake at night wondering about how the results will turn out in the end. Secondly, the way you write the thesis depends to a large extent on what you find in the empirical analysis. Thirdly, peeking at the results allows you to start writing at an early stage, which is important if your aim is to be finished before the deadline.

Most hypotheses regarding the relationship between variables can be cast in terms of a regression model, which makes ordinary least squares (OLS) a natural estimator. Although more advanced techniques are certainly possible, you will probably feel more comfortable with OLS. This is therefore a good estimator to consider when trying to get a feeling for the data and what kind of results you might expect to end up with. Once you know how the results look, you are in a much better position to decide if it is going to be worthwhile to pursue more advanced econometric techniques. It also gives you an indication as to whether you have all the data, or if you need to look for more. Maybe the sample is just not large enough for reasonable precision in the estimates, and that this is why the results look a little strange? Or is it maybe that you have omitted an important variable that simply must be there?

In sum, then, the writing process starts with a rough idea of a question and an hypothesis, which is then subjected to a preliminary test using either the complete data set or just a subset. With these results in mind one then goes back to square one again, and rework everything properly, making sure that the question is relevant, that the data set is complete, and so on.

3 Structuring the thesis

The purpose of the written thesis is not just to report the results of your research but to present them in a way that makes the reader interested. Before reading the thesis the reader does not know whether he or she is going to enjoy it. It therefore falls on you to tell the reader that this is indeed interesting enough to spend time on.

Most of the material that you will read during your search for a good research question is going to be journal articles, which are condensed to conform with the journal's requirements regarding the number of pages. The working paper versions of these articles are usually much longer and also more detailed. In this sense, a Master thesis is more like a working paper.

However, the structure is basically the same as that of a journal article. Use sections and subsections to partition the material. For example, if Section 6 is "Empirical results," then you may want to have a subsection, Section 6.1, called "Data." But do not get carried away; two section levels are enough. With three or more levels the text looses much of its flow, and the writing tend to become more mechanical. The following format represents a minimum requirement:

1. Title page.

2. Abstract.

3. Acknowledgements.

4. Introduction and literature survey.

5. Theoretical analysis.

6. Empirical analysis.

7. Conclusions.

8. References.

9. Appendices.

Readers are busy and impatient. No reader will ever read the whole thesis from start to finish. Readers skim. You have to make it easy for them to skim. Most readers only want to know your basic result. It is therefore customary to organize the thesis in "newspaper" style (Cochrane, 2005). Notice how newspapers start with the most important part, filling in the background later for those readers who kept going and want more details. This is exactly the purpose of the above format. The abstract works as the "headline" and explains the main contribution, which is then explained in the main text. By contrast, a good mystery novel has a long windup to the final punchline, during which time the reader is completely unaware of what to come.

Keep the thesis as short as possible. Final theses should be no more than 25 pages for single authors, and 50 pages if you are two or more. Shorter is usually better.

3.1 Title page

The title page should contain the title, the date, and the names and personal numbers of all the authors. The title should be short and informative. Just by looking at the title the reader should get a good idea of what the paper is about. But it should not reveal too much – it should be just enough to make the reader curious.

3.2 Abstract

The abstract should be short, about 100–150 words. The main function of the abstract is to communicate the one central and novel contribution. Write this down in one paragraph. As with all your writing, this must be concrete. Do not write "I analyzed data on exchange rates and found many interesting results." Explain what the central results are. To take an extreme example, in their abstract Dickey and Fuller (1979) write:

> "Let n observations $Y_1, ..., Y_n$ be generated by the model $Y_t = \rho Y_{t-1} + e_t$, where Y_0 is a fixed constant and $\{e_t\}_{t=1}^n$ is a sequence of independent normal random variables with mean 0 and variance σ^2. Properties of the regression estimator of ρ are obtained under the assumption that $\rho = \pm 1$. Representations for the limit distributions of the estimator of ρ and of the regression t test are derived. The estimator of ρ and of the regression t test furnish methods of testing the hypothesis that $\rho = 1$."

Given some basic econometric knowledge, this abstract does everything that it is supposed to; it is short, concrete and communicates the main contribution. Other literature should not be mentioned in the abstract. Here is another example, taken from Miguel et al. (2004):

> "Estimating the impact of economic conditions on the likelihood of civil conflict is difficult because of endogeneity and omitted variable bias. We use rainfall variation as an instrumental variable for economic growth in 41 African countries during 1981–99. Growth is strongly negatively related to civil conflict: a negative growth shock of five percentage points increases the likelihood of conflict by one-half the following year. We attempt to rule out other channels through which

rainfall may affect conflict. Surprisingly, the impact of growth shocks on conflict is not significantly different in richer, more democratic, or more ethnically diverse countries."

3.3 Acknowledgements

It is customary in theses to acknowledge any substantial help. Usual suspects include family, fellow students and thesis advisors.

3.4 Introduction

The introduction should motivate and describe what you do in the thesis. As with the abstract, the introduction should be concrete and to the point. Some students do not know what to include and therefore tend to become very handwaving in their writing. If you do not know how to structure the introduction, take a look at an article you like and adopt a similar approach.

Your readers are busy and impatient, and unless you are able to pursue him or her to continue, most will stop at the introduction. A good motivation is therefore absolutely crucial! A common way to start is to begin with a brief description of the problem under consideration, which is followed by the motivation. As mentioned earlier, most readers do not know that they are interested, so it is up to you to tell them. Maybe the evidence has been mixed, or the evidence has been clear but the interpretation has not? It is controversies like this that makes the question interesting. The next step in the motivation is to explain how your thesis comes into the picture and your central contribution. The motivation should therefore address three issues:

- What is the nature of the problem under investigation?

- Why is it interesting?

- What does your thesis attempt to do? That is, what is the contribution of your thesis?

Once the motivation has been taken care of, the next step is to describe in more detail what you do in the thesis. As with all your writing, this must be concrete. You must write so that people can understand how the results came about. Do not just state your conclusion: "My results show that the PPP hypothesis has to be rejected." This is what the abstract is for. Give the working behind the result: "In a regression of Y on X, controlling for Z, the

estimated slope coefficient is significantly different from zero, suggesting that PPP must be rejected."

Of course, doing it the other way around works just as well. Then you begin by stating your main findings, and add the motivation later. In any event, it is good to wait a while with writing the introduction. The reason for this is that the writing becomes much easier once you know what your results are. A good rule is to at least write down the empirical results before you begin with the introduction.

It is customary to end the introduction with a brief roadmap paragraph: "Section 2 sets out the model, Section 3 gives the main results, Section 4 checks for robustness," and so on. Three pages is a good upper limit for the introduction.

3.5   Literature review

The thesis should be well-positioned relative to the rest of the literature. However, it is not necessary to have a separate review section; oftentimes it more natural to have the review in the introduction, when you motivate the problem. Of course, readers are mostly interested in what you do. So, if the review is long, it might be a good idea to have a stand-alone section that people can skip if they are not interested. But then it is useful to have some of the key references also in the introduction, since otherwise it may be hard for people to understand how your thesis is different from others'.

If the study you referring to is important it is common to name it directly in the text: "Dickey and Fuller (1979) were among the first to consider the limit distribution of the estimator of ρ when $\rho = \pm 1$." On the other hand, if the reference is less important, then you may put it in parentheses at the end of the sentence: "The results of previous studies are mixed, and not very convincing (Westerlund, 2008)." If there are more than two authors, only the first author should be used, followed by "et al." meaning "and others": "Larsson et al. (2009) find that the PPP hypothesis must be rejected." In the text it is therefore only the names of the authors plus the year of publication that is given. There are many reference styles around but this one, the Harvard reference system, is the most common in the field of economics.[2]

The literature review is a summary of the major studies that has been done in your area. However, it is not necessary to cite every single paper that you can find. The main point is to

---

[2]See Section 3.9 for a further discussion.

set your thesis off against the two or three closest previous studies, and to give proper credit to people who have come up with things that might otherwise seem new in your thesis. According to Greenlaw (2006) the literature review has three goals:

1. It should identify and explain the main findings of the two–three closest studies.

2. It should point out the main deficiencies of these studies.

3. It should explain how your thesis contribute by overcoming the deficiencies of the previous studies.

That is, the main function of the review is to guide the reader through the literature, and in so doing to selectively point to those weaknesses that you will address in your thesis.

Remember that the literature review is just review, which should be kept as short as possible. Some researchers write long reviews that are followed by a short analysis of the results, which is understandable since much time goes into reading articles. Unfortunately, the review cannot be considered as a contribution of your thesis, and it is therefore better to save your efforts until the empirical analysis.

3.6 Theoretical analysis

The purpose of the theoretical analysis is to present the theory behind the problem you are investigating. In other words, the purpose here is mainly to help understand and to put into context the model you take to the data. In most, if not all, theses the theoretical model is just an equation, whose estimation is then the main concern of the empirical analysis. Your job in the theoretical analysis is therefore to derive the equation in question and, perhaps more importantly, to derive testable hypotheses. In a regression equation setting, this amounts to stating the expected sign and magnitude of the regression coefficients.

But even if you are just using theory to explain the equation that you are going to estimate, the theoretical analysis is still important. The reason is that the validity of the empirical analysis depends on how reasonable your theoretical model is. If the theoretical arguments are weak, then the empirical results will be weak, too. In fact, many of the assumptions you make are not even testable.

Causality is one example of such an assumption, which cannot be inferred from the data. One of the purposes of the theoretical analysis is therefore to establish which variable that is

going to act as the dependent variable and which ones that are going to as explanatory variables. Explain how the causality works in economic terms. Are you sure that the causality runs from X to Y, and not from Y to X, or from Z to Y and X simultaneously? If the direction of the causality is not clear, so that the error term is potentially correlated with the regressors, then explain the economics of why your instruments are correlated with the regressors and not with the error term.

Economic theory can also be used as a guide to decide which control variables to include. Each variable should have its own little "story." Choice of functional form is another example where economic theory might be useful. For example, if you would like to explain the relationship between a firm's profit and the price it charges, then do not just settle with a simple regression of profits on price, because this how the linear regression model is usually presented in textbooks. Adapt the model so that it makes more economic sense; the relationship between profits and price is more likely to be concave than linear, suggesting that price squared should be included as an additional regressor.

The theoretical analysis should be the minimum required for the reader to understand the empirical results. Do not write a up a general model and then "for the empirical work, we now specialize the general vector autoregression of order p to a simple autoregression of order one, we use only two firms rather than a continuum, we assume agents have quadratic utility," and so on. Work out only the specialized model that you actually take with you to the data. You may even adapt the notation to fit the problem at hand. For example, if the dependent variable is per-capita GDP at time t, write "GDP_t;" it is not necessary to always refer to the dependent variable as "Y_t."

3.7 Empirical analysis

The purpose of the empirical analysis is to provide empirical evidence for your research hypothesis. Except possibly for the introduction, which is where you persuade the reader to continue reading, this is the most important part of the thesis. The empirical analysis should include the following elements:

- A description of the data used.

- A discussion of the empirical model.

- A discussion of the econometric analysis used.

- An explanation of the empirical results and how they compare to the hypothesized ones.

The empirical analysis is the main contribution of your thesis, and should take up most of the space. However, it should still be focused. Try to sort out the results that are central to your thesis. This will requires some thought. It will also cause some pain, because you will start to realize how much you are going to have to throw out. Once you have done it, however, you are in a much better position to focus the thesis on the question of interest.

Do not do extensive warmup exercises, although some data description and preliminary estimation results should always be included. You may for example illustrate the main features of the data using graphs and simple descriptive statistics, such as means, standard deviations and correlations. Also, try not to motivate the specification that worked with your failures. Many students start off by analyzing a too simple model, which later on ends up being rejected by the data. A more general model is therefore proposed, and the same analysis is done all over again. But by then the reader will have lost much of his or her patience. Busy readers do not have time to hang around for multiple rounds of analysis just because you have not decided on which model to use. If any of these results are really important, they can come afterwards, as a robustness check.

If you have found the best model but still want to show some of your early attempts, then it is a good practice to report the results from the different models alongside each other. This might be interesting from a robustness point of view, or because you would like to illustrate something that does not require the extra complexity that comes from considering the fully specified model.[3] The main difference is that now the reader knows that you are aware of the fact that the simpler model is not the correct one, and therefore that one should be careful not to overstate the results.

Similarly, although probit may stand out as the most appropriate econometric technique to use, you may still want to report the results from OLS. In particular, while best from a theoretical point of view, probit is also more complicated. OLS has the advantage of being simple, which makes it a natural benchmark.

Once you have presented the results you need to interpret them to the reader. How does the estimated coefficients compare with the ones predicted by theory, and how do they com-

---

[3]This is Occam's razor, a well-known corollary to the scientific method. It states that when choosing among a alternative theories to explain a particular phenomenon, one should select the theory with least complexity.

pare with the results reported by others? Apart from the actual comparison of the results, reading other studies is very rewarding because it gives you new ideas of how to make the interpretation of the findings richer. Also, are the estimated coefficients statistically significant? Are they economically significant? It is perfectly possible that, while statistically significant, the estimated coefficients are still too small to make a difference from an economic point of view. Any anomalies should be carefully discussed and at least partially explained.

If possible the main results should be followed with graphs and tables that give intuition. Follow that with some responses to obvious criticisms and robustness checks.

Some students have been taught that in the empirical analysis one is not permitted to draw any conclusions about the results, and that one should simply let "the data speak for themselves." This is not the case, at least not within the social sciences. One could even say that it is the other way around. That is, in economics you are expected to have a lot to say about the results!

Finally, remember that a failed hypothesis does not imply a failed project! In fact, this is the whole point of the preliminary testing; it allows you to prepare and to write up the text to fit the results. Suppose that you have decided to test the Keynesian theory of consumption, which proposes that consumption is largely determined by income. But your regression results suggest that income is insignificant. Then it is perfectly fine to conclude that the hypothesis of Keynesian consumption fails. In fact, not finding any support of a well-known theory is in some sense more fun than simply showing that it works, which is what people would expect.

3.8 Conclusions

A concluding section should not be necessary. If you did a good job of explaining your contribution in an understandable way in the abstract and introduction, and then documenting those claims in the empirical analysis, then there is really no need for any conclusions (Cochrane, 2005). The conclusion should therefore be kept short. It should also not be repetitive; do not cut and paste what you wrote in the abstract or introduction. One suggestion of how to bring some substance into the conclusions is to have a paragraph acknowledging limitations. You can also use the conclusions to discuss implications for policy, and so on.

3.9 References

As mentioned earlier there are many reference styles around, which all have their own way of making scholarly reference. The Harvard reference system is probably the most commonly used style within the field of economics. In this style, references in the text are presented as explained in Section 3.5.

The proper way to write references in the reference list depends on the type of publication. The required elements for making reference to a book are: Author, Year. *Title of book*. Edition (only include this if not the first edition). Place of publication: Publisher. For example,

> "Gouriéroux, C., and Monfort, A., 1996. *Simulation-Based Econometric Methods*. Oxford, UK: Oxford University Press."

For journal articles the required elements are: Author, Initials., Year. Title of article. *Full Title of Journal*, Volume number (Issue/Part number), Page numbers. For example,

> "Bai, J., and Ng, S., 2004. A Panic Attack on Unit Roots and Cointegration. *Econometrica* 72(4), pp. 1127–1177."

The Harvard system does not include any guidelines for working papers. However, a common convention is to use the following elements: Author, Initials., Year. Title of article. Type of working paper. For example,

> "Levin, A., and Lin, C., 1992. Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. University of California working paper 92-23."

Sometimes the paper does not have its own working paper number, in which case it is common to just label the paper as "unpublished manuscript." Publications available on the Internet, newspaper articles, dissertations, and so on, also have their special requirements, which are easily found via Google. The reference list should start with a new page.

3.10 Appendices

Researches often want to document that the main result is robust to various other ways of doing things. You should also do that, but once you have found the one best way of doing things, it is not worth the space to present all the checks and variations, at least not in the

main body of the thesis. Appendices are very nice in this regard, as a means to store away results that are of "secondary nature."

However, you should be careful not to place too much in the appendix. The general rule is that if a result is important enough to include then it should be in the main body. In fact, most journal articles does not have any appendix at all. But if you do decide to place some of the material in an appendix, remember to mention this in the text, so that interested readers know that there is more. As with the reference list, the appendix always starts with a new page.

4   Getting it down on paper

Most researchers spend at least 50% of the time they put into any project on writing. A well-written text with carefully prepared figures and tables provide a sense of quality, even if the research question is totally uninteresting and the data poor.

Pay attention to the writing in articles you read, and notice the style adopted by authors you like. Here are a couple of other things to think of when preparing the manuscript.

4.1   Reporting empirical results

- A good rule of thumb is that first you describe what you do, then you explain it, compare it to alternatives, and compare it to the procedures of others (Cochrane, 2005). For example, in describing a data transformation, start with saying: "All variables are expressed in logs." Then explain why taking logs is important, and if there are any other alternative transformations that fulfill the same purpose.

- Document your work. A fellow student must be able to sit down with your thesis and reproduce every number in it from the instructions given in the text. Even the typical research paper falls short here. It is not uncommon to find that the description is not enough to figure out how the table of results was computed, how the standard errors were computed, how the estimation was carried out, etc.

- Try to keep track of what your reader knows and does not know. Most researchers assume far too much. What the reader usually wants most of all is to understand your basic point. If the road there is long and all steps are not clear, there is no chance that

the reader will be able to grasp the main result. It is therefore better to assume too little rather than too much.

4.2 Tables

- Each table should have a self-contained caption so that a skimming reader can understand the facts presented without having to go searching through the text for things like the definitions of Greek letters. But do not overdo it; details of variable construction and such things can be left out. The goal is to allow a skimming reader to understand the table, not to substitute for the detailed documentation that should be provided in the text.

- No number should appear in a table that is not discussed in the text. You do not have to mention each number separately; "Row 1 of Table 3 shows that income is insignificant in all three regressions considered" is okay. But "Table 5 shows summary statistics" is not okay. If it is not worth discussing in the text, it is not worth putting in the table.

- Never copy directly the tables produced by the software program! Paste it into Excel first and strip it from everything that you do not need, and make sure that you use the correct number of significant digits, not whatever the program spits out. 4.5678356 with a standard error of 0.6789345 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. Also, try to use sensible units. Percentages are good. If you can report a number as 2.3 rather than 0.0000023, that is usually easier to understand.

- Tables take a lot of space – use them efficiently. If you run two regressions, then do not report the results in two separate tables; make one table containing the results from both regressions. Besides saving a lot of space, merging results in this way makes comparisons simple. You may also want to include R-square and other relevant test statistics for possible econometric problems like heteroskedasticity and autocorrelation.

- Never report results that you do not know what they mean or intend to discuss in the text. For example, if you are interested in the significance of the intercept, then there no need to report the t-statistic, the 95% confidence interval, the standard error,

and the p-value; the last two are enough. Similarly, if the STATA regression output is complemented with "robust standard errors," then you need to be sure that you understand exactly what this means before you include it in the text.

4.3 Figures

Good figures can make a thesis come alive, and they illustrate patterns in the data much better than big tables of numbers. Figures are therefore perfectly suited for preliminary analysis, but can also be used for illustrating a main result. Whatever you intend to show later in your analysis should be visible in the data; if you argue that there is a linear relationship between X and Y, then the observations should tend to lie close to an imaginary straight line passing through the middle of the data. If the tendency is weak, chances are that the regression results are going to be weak, too.

But figures are good not only for showing things to the reader, but also when trying to understand things yourself. Students oftentimes get so ensnared in the technicalities that they forget about looking at the data. Figures and common sense are usually just as important as test statistics and p-values.

Give a self-contained caption, including a verbal definition of each symbol in the figures. Label the axes. Use sensible units.

4.4 Footnotes

Do not use footnotes for parenthetical comments. If it is important, put it in the text. If it is not important, delete it. Obviously, a lot of parentheses are just as bad as a lot of footnotes. Use footnotes only for things that the typical reader 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.

4.5 Other writing tips

- Simple is better. Researchers and students alike sometimes think they have to dress up a paper or a thesis to look impressive. The exact opposite is true; the less math used, the better. The simpler the estimation technique, the better.

- Do not use difficult words just to impress. Use simple sentences that readers can understand, and avoid technical jargon wherever possible. Use simple short words, not big

fancy words; "use" not "utilize," "several" not "diverse," etc. You should also avoid using adjectives to describe your work; "striking results" and "extremely significant" are not okay.

- Write with a few words as possible. As you edit the thesis ask yourself constantly: "Do I really have to say this, and if so can I make the same point in less space?" Do not repeat things; if you have said it once, there is usually no need to say it again. It uses up extra space and readers' patience to have to see the same point made over and over again.

- Present tense is usually best. You may say "Dickey and Fuller (1979) find that" even though 1979 was a while ago. Similarly, it is not necessary to write in the future tense when you are talking about things to come. It is for example perfectly fine to write in the abstract that "in this paper we test PPP" even though the actual test has not been presented yet. Most importantly, though, keep the tense consistent; do not start a paragraph in past tense and finish it in the future.

- It is not necessary to use "we" in a sole-authored thesis. A common practice is to use "we" when you refer to both the reader and yourself, and "I" when referring to yourself only. For example, "as we can see in Table 5 none of the coefficients are significant" and "I used OLS to estimate the regression" is fine.

- Use the active voice. Researchers tend to use the passive voice, in which the subject does not performs the action. For example, instead of "I analyzed the data" they write "the data were analyzed." Similarly, instead of just writing "we argue that" they write "it is argued that." The active voice is usually clearer.

- Do not let sentences and paragraphs become too long. Write short paragraphs. Reading is easier when the text is not jammed together. In general, a paragraph should make one point, or a part of a broader point. Readers therefore expect a paragraph to be short and with a single focus.

- Make sure that the grammar is correct. Given the spell-checking capabilities of most text processing software packages like Microsoft Word, editing the text is relatively easy.

- Readers forget Greek letter definitions. If you define δ once in an obscure part of the text and then write "$\delta = 3$ gives the best fit," no one will know what you are talking about. Define these letters clearly in an easy-to-find place, preferable in an equation that is set out from the text, or in connection to one. It is best to give them a name too, and then remind readers of both the name and the letter. For example, "I find the best fit when the elasticity of substitution δ equals 3."

- Shorthand notations such as "OLS" should be written out in full with the shorthand in parenthesis on the first occasion. Thereafter it is enough to use only the shorthand. If the word – here "ordinary least squares" – is used only on a couple of occasions, there is hardly no need for a shorthand.

5 Conclusion

This paper provides some writing tips for theses in applied econometrics. One of the most important lessons is to start early, and to begin looking for data as soon as the research question becomes clear. If the question is interesting and the data are suitable, the thesis project is likely to be a success!

References

Cochrane, J. H., 2005. Writing Tips for Ph. D. Students. Unpublished manuscript.

Dickey, D. A., and Fuller, W. A., 1979. Distribution of the Estimator for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74(366), pp. 427–431.

Greenlaw, S. A., 2006. *Doing Economics: A Guide to Understanding and Carrying out Economic Research.* Boston, USA: Houghton Mifflin Company.

Miguel, E., Satyanath, S., and Sergenti, E., 2004. Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy* 112(4), pp. 725–753.