

# QMDA Project Advice

## 1. Introduction

You will find that doing the project helps you understand the theoretical material and pays off in better performance in the exam. Being able to show potential employers a well presented piece of empirical research often helps get a job. The techniques that are used in the project can greatly help you for your MSc dissertation, if you wish. These notes provide advice on doing the project, ignoring this advice may result in you being penalized when the project is marked.

During this course you must learn how to produce and interpret a regression output, by doing the applied exercise. Try to write up what you are doing as you go along and build in an 'audit trail'. It is very easy to make mistakes and to forget what you did, having records and documentation is essential. Make notes of exact references when you read articles, trying to search for the reference at the last moment is a waste of time. Date your drafts so you can identify them and do not submit an old version by mistake.

Furthermore, make sure that you keep multiple back-ups of your data and drafts (e.g. on your hard drive and a USB stick). You can lose computer files in lots of different ways and it is a lot of work redoing everything. I hear about a lot of computer disasters as the deadline approaches and I will not be sympathetic if you are having trouble completing because you did not back-up properly.

## 2. Rules

- Submit the project by June 11 2020.
- If you do not submit by the deadline the usual penalty scheme of the school applies
- The maximum length of the project is 10 pages, Tables, graphs and technical appendices are not included in this total.
- The data you used should readily available to me in E-views file in a way that allows me to replicate your results.
- The first page of the project should have;
  - Title,
  - Abstract,
  - Name,

- The programme you are following
- The pages should have page numbers.
  - The project should be divided into properly titled sections, with an introduction and conclusion.
  - The project should have a clear description of the data: definitions, units of measurement, etc. The data should be appropriately graphed.
  - The project should look and read like a very professional business report.
  - Do discuss the project with other students, people at work, etc. But the project must be your own work. Plagiarism is heavily penalized and it is more difficult to get away with than you may think. If in doubt, ask advice from the staff. I interview a selection of students to check that their project is their own work.

### **3. Stages of doing your project**

#### **3.1. Getting to know your data.**

You do not have to get a “Good” model. A “Good” is often a matter of luck. Do not be surprised if you cannot get a good model (right signs, passing misspecification tests) for a standard equation. The published results are often based on a lot of data mining (experimenting with different samples, estimation methods, variables including dummy variables) and the ones that do not work even after all this do not get published.

Do not paste program output into the text of the project; summarize it properly in equations or tables. Program output should not be included. Read empirical articles and try to copy their style and approach. I do not assess you on the product, the quality of the final model, but on the process. I do this under five categories: data, theory, econometric analysis, presentation, written style. There is some substitution, e.g. if you put a lot of work into some specific parts of the project, I will give you extra credit for that, but there are diminishing marginal returns, so make sure you put effort into all elements. Remember that when I mark the project, I have the data you used. It is very easy to check what you did and what you missed. I do check.

#### **3.2. Develop a Theory**

Theory should be interpreted very widely here: what do we know about the process that might have generated the data? Standard theory may tell you what variables are likely to be relevant: in a demand function: income and own and other prices will appear on the right hand side. Theory may tell you about functional form, but usually does not. If the variables used are always positive start with logarithmic transformations.

It is often quite difficult to translate pure economic theory into the form of an equation that can be estimated, but wider theory is often useful

in giving you a starting point. Furthermore, a central issue is distinguishing correlation from causality. Theory should help with this distinction. Sometimes you do not want to make causal statements, just forecast. Use your common sense to develop the theory, ask: is this a sensible way of explaining the data, and try to identify the interesting questions.

### **3.3. Estimate the specification**

Your examination of the data and review of the theory should have given you some ideas about designing the models that you will estimate: the variables that you include, the functional form that you use, questions that you need to answer with hypothesis tests, the sign and magnitude of the coefficients you expect.

You must organize the estimation process. It is very easy to make mistakes; being organized makes this less likely. It is very easy to lose files: make back-ups on separate disks and establish a system for naming files and variables. It is very easy to get buried in vast piles of regression output: organize your estimation to stop this happening. Getting lost is particularly easy when for each regression you also calculate diagnostic tests for normality, misspecification etc..

Look at the magnitude of the coefficients are they sensible? What values would you expect from theory? An effect may be statistically significant, but so small as to be economically unimportant. Or large in economic terms, but imprecisely estimated so not statistically significant. Remember that our conventional significance levels, like 5% are just conventions, other levels may be appropriate.

The first stage in getting organized is to write up as you go along. Before you start estimating anything you should have written the first draft of the data analysis and theory sections, with predictions for likely values of the estimated coefficients. The second stage of getting organized is to be systematic in your estimation. Go to the computer with a plan of what you are going to do, organized around the crucial questions and a draft table to summarize the results. If you have designed a table, you can just put a cross in the box (or the  $p$ -value) if it fails the normality test for instance.

## **4. Write up the results**

The final project should read like a business report, not a part of your *autobiography*. All the problems you had (finding the data; making the computer work; trying to understand the articles; your personal crises) do not belong in the project. Read academic articles to see how they present their results, but bear in mind that the way research is presented in the

literature is not a good guide to how it was actually carried out. Do deals with fellow students to read drafts of each other's projects to see if they are clear and to correct errors. Worry about spelling, grammar, construction of sentences and paragraphs and try to make the writing lively and interesting.

*Introduction.* This should motivate the questions you are asking, provide some background and explain why the issue is interesting.

*Theory.* Provide a brief link to the literature, set up the model and any hypotheses you want to test. Set out the questions you are going to try and answer.

*Data.* Give exact definitions, units and sources; discuss any measurement problems, how the variables relate to the theoretical concepts; characterize the basic properties of the data, identify any trends, seasonals, cycles, outliers *etc.* You must give graphs which summarize the main features of the data. If you miss something which would be very obvious had you graphed the data I will penalize you. You may want to include a separate data appendix with more detail.

*Statistical Model.* Present your specification and briefly discuss the estimation method you are going to use and why it is appropriate. This involves justifying the assumptions that you made about the distribution of the error terms. In most cases you will use Ordinary Least Squares. Do not put text-book material in your project, e.g. proofs that OLS is BLUE. Just give a reference to the text-book. The project should contain all information that we do not know but need to know, not things we know.

*Results.* Make sure that you organize the presentation of your results clearly, bringing out the important ones and referring to the others in passing. For instance, if you tried a variable that proved insignificant, just say in a footnote that you tried it and it proved insignificant, you do not have to go into detail and to present the entire specification. Think carefully about how you want to present the numerical results, either as equations in the text or tables. What information do you need to convey? This will include the estimated coefficients, their  $p$ -values (or their  $t$ -ratios, but only one of the two). Make sure you explain your results, e.g. say whether you give standard errors or  $t$ -ratios. Look at empirical scientific articles and see how they convey the information. You don't need to put program (E-Views.) output as an appendix, but the main text should convey (creating your own Table) the crucial results in a comprehensible form. Make sure that you interpret the results in terms of the substantive issues and consider both the size of coefficients and their significance.

*Conclusions.* What did we learn from this project? How were the questions posed earlier answered? What is their relevance for practical questions of forecasting, policy etc? Are the answers consistent with theory or institutional information? Is the model statistically well specified?

*References.* Make sure you follow a standard style of referencing (e.g. Harvard style or Chicago Style). I prefer the Harvard style ☺.

*Appendices.* More detailed output or technical derivations which are not in standard sources can be put as appendices.

## 5. General advice

Read, write and think. Keep reading and relate your problem to what is in the literature. Read more applied articles and see how the authors did it. Make your project look professional, something that might get published. Potential employers often ask to see projects, so keep a copy for yourself. Try and follow the examples of professional writing in the literature. Start writing early and keep rewriting. Organize your empirical investigation, so you do not get lost. Back up your computer files. Focus on particular questions. Explain why these are interesting questions. Try to make the project clear, brief and interesting. Make sure you have got the references right, make sure that when you quote from another paper you put it in quotation marks and give the exact source. Keep rewriting it to achieve those goals. Get other students to read it and comment on it. Remember you are not being marked on how good the final model is, you are being marked on how you went about it and how you reported what you did. You will be penalized if you have not taken account of advice in these notes. Enjoy the process; it can be fun discovering new things.

*Always Deliver more than Expected*