Summer 2024 **Harry J. Paarsch**

University of Central Florida
Department of Economics

ECO 6935
Capstone in Business Analytics I

Completing the Capstone Project Successfully

These notes are intended to provide you with some guidance in deciding on a topic for your capstone project, completing the modeling and data gathering required, and subsequently writing up the results of your efforts.

1. Choosing a Topic

No capstone project can ever be successful without an interesting problem, but some interesting questions can take several years to answer. Therefore, strike a balance between interesting and feasible topics. During the first few weeks of the course, you should be actively thinking about a potential project. The topics you consider should not only be of interest to you, but also have some business relevance. Try to focus on a well-defined problem. Because it is far too broad, the following is neither:

How can firms eliminate waste in businesses?

Answering this question would involve solving many, many problems. On the other hand,

How much does a Apple MacBook Pro cost?

is a narrow and uninteresting question. A question I once posed as the topic of a paper is

How to score past customers to decide on follow-up e-mailings?

In all cases, before you embark on a project, consult with me through your written project description.

By 27 May 2024, you must submit to me a written description of the project you wish to undertake. In this document, you should outline why the topic is of interest to you and why it is of business relevance. "Because" is an insufficient answer to either question. You should provide a brief description concerning how you propose to answer the question posed, too. Limit yourself to one $8\frac{1}{2} \times 11$ inch double-spaced, type-written page. I shall read this document and write suggestions on it. Should your topic be unsatisfactory, then you will subsequently meet with me to discuss possible changes, or an alternative topic. From your interests, we should be able to define an appropriate, feasible problem—one that can be investigated in two months.

Central to success in this part of the project is that you not procrastinate: few have either the creativity or the luck to think up a good project overnight. You should take this exercise seriously; successfully completing this capstone project is an important

accomplishment—one valued by employers. Moreover, a successful capstone project can become part of your resumé and an important part of your future. The investments you make while undertaking this capstone project will pay dividends later. In any case, you cannot receive credit for this course without having submitted on time an acceptable topic and an acceptable outline; you cannot earn your master's degree without having successfully carried out your capstone project outline, on time.

2. Creating an Outline

To conduct your research in an orderly manner, you should have an outline. This outline is a map of how you will proceed in solving the problem you have posed—a blueprint for you to follow when completing your empirical work and writing up your results in the next course, ECO 6936. You must submit the outline before 17 June 2024.

A useful outline has several parts: One is a list of background readings.[1] Another is a theoretical model of the phenomenon you are interested in investigating. (This is perhaps the most demanding part of the project and it will require the most creativity.) Yet a third part is finding and documenting available data sources.

2.1. Developing a Reading List

After having chosen a topic, your next step is to find out what other researchers have done with this sort of problem, or with another similar problem. This will involve searching the literature related to the topic. Such search can occur in at least three potential places.

You might think that the Internet would be a natural place to search, but you would be mistaken. For even though the Internet has many virtues, one problem with it is much of the material on the Internet is of dubious quality. Instead, you should start with a database compiled by the American Economic Association, named *EconLit*, in which most published and many unpublished economics papers are organized and can be searched by keyword(s). This is perhaps the most promising source available in the profession.

Another potential source of information is the *Journal of Economic Literature*, which used to be called the *Journal of Economic Abstracts*. This journal, often referred to as JEL, publishes surveys on topics of interest to all economists as well as lists of paper titles by subject area. The surveys are typically accompanied by extensive bibliographies and these papers can save you hours of searching.

Yet a third source is the *Web of Science* (WoS). This online database contains lists of published papers in which citations made to the work of another author are tabulated and documented. (WoS is superior to Google Scholar because the latter lists all the papers it can find on the Internet, which can make it difficult for you to discern quality.) In order to use the WoS successfully, however, you must have a seminal article. By searching the WoS, you can find papers which criticize, evaluate, or contribute to the seminal paper.

If, after some searching, you have been unable to find these sources in the library, then ask one of the librarians for assistance. If, after considering all of these sources, you

[1] In order for you to know whether these readings will be of use, you will actually have to read them before submitting the outline.

have been unsuccessful in developing a useful list of readings, then you should ask me for assistance. I am here to help you, but you should not use me as a substitute for your work. One purpose of this project is to make you familiar with the resources available, and be comfortable in using them.

Caution: do not spend all of your time compiling the perfect list of readings; select a few (good) ones and go from there. Be sure that data sources that complement the literature you are reading (see section 2.3) exist, too.

As you compile a list of readings, maintain a database of bibliographic information. Because you will be required to use the reference management software BibTeX when formatting the references for your submitted paper, your best option is to create a `References.bib` file as described in Chapter 11 of the book *A Gentle Introduction to Effective Computing in Quantitative Research: What Every Research Assistant Should Know*, by Konstantin Golyaev and myself.

This information will be essential when you submit your proposal. That said, you should realize that many different styles exist. For the purposes of this proposal, however, you should adopt the following guidelines:

`http://www.chicagomanualofstyle.org/home.html`

The above has more information than you will require right now, so let me illustrate briefly how several types of documents should appear in the list of readings. Note, too, that the file `chicago.bst` is a BibTeX style file that can help you in this regard.

The first document I want to consider is a book. The appropriate bibliographic citation is

> Lastname, Firstname (Year). *Title*. Place of Publication: Publisher's Name.

For example,

> Koopmans, Tjalling. (1957). *Three Essays on the State of Economic Science*. New York: McGraw-Hill.

If more than one author exists, then the citation should be

> Arrow, Kenneth and Frank Hahn (1971). *General Competitive Analysis*. San Francisco: Holden-Day.

When more than one edition of the book exists, list the edition used. For example,

> Wonnacott, Thomas and Ronald Wonnacott (1984). *Introductory Statistics for Business and Economics*. Third edition. Toronto: John Wiley & Sons.

The next kind of document is a journal article. The convention concerning one or several authors is adopted here, too. Unlike a book, however, additional information is required. Not only must you list the title of the article and the name of the journal, but you must also list the volume of the journal, the year, and the relevant page numbers. For example,

> Lastname, Firstname (Year). "Title." *Journal* Volume, beginning page–ending page.

Thus,

Hendry, David and Jean-François Richard (1983). "The Econometric Analysis of Economic Time Series." *International Statistical Review* 51, 111–163.

The third type of document is an article in a book edited by another author. Again, the convention concerning the order of names is maintained, but now, in addition to the title of the article, you must also list the title of the book in which the article appears, the editor of the book, and the relevant page numbers as well as the publication information given for books. For example,

Lastname, Firstname (Year). "Title." In Book Editors (Eds.). *Book Title*, Volume, beginning page–ending page. Place of Publication: Publisher's Name.

Therefore,

Engle, Robert F. (1984). "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics." In Zvi Griliches and Michael Intriligator (Eds.) *Handbook of Econometrics*, Volume II, pp. 775–826. New York: North-Holland.

One, perhaps unusual, type of document is a working paper or manuscript prepared by a researcher at a university or research institute. In this case, the citation is

Lastname, Firstname (Year). "Title." Working Paper #, Institution.

Hence,

Gunderson, Morley, John Kervin, and Frank Reid (1985). "Effect of Labour Relations Legislation on Strike Incidence," typescript, Department of Economics, University of Toronto.

Another document you will likely need to reference is a government publication. In this case, you should use the following format:

Government Agency (Year). *Title*, catalogue number.

For example,

Statistics Canada (1983). *Production, Shipments, and Stocks on Hand of Sawmills in British Columbia*, catalogue number 35–003.

The final type of document you will likely need to reference is a web page. In this case, you should use the following format:

Last Name, First Name. "Page Title." Website Title. Web Address (accessed Date Retrieved).

An instance of which is

Foroux, Darius. "How To Stop Wasting Time and Improve Your Personal Effectiveness." DariusForoux.com. http://dariusforoux.com/stop-wasting-time/ (accessed March 9, 2019).

2.2. Constructing a Model

Modeling a real-world phenomenon is difficult: it requires experience. Most of you have had none of this, but your capstone project can change that. To gain experience, one strategy adopted by many researchers is simply to borrow models which others have created. Of

course, such a strategy has risks: the model you choose need not be applicable. In other words, if you choose to borrow a model, then be sure it applies. In any case, if you do borrow some one else's model or imitate the work of others, then be sure to acknowledge this debt. (I shall discuss this at length in section 3.4.) Some successful projects do not involve any economic modeling whatsoever, but entail instead measuring the effect some event or policy has had on a business, but even such exercises involve some modeling, such as statistical modeling.

2.3. Finding Data Sources

Finding data sources that are well-suited to the needs of your research can be difficult. In fact, whether the data are available to answer the question you have posed will typically be the criterion determining whether your topic is feasible. Clearly, the sources you consider will depend on the problem you have chosen. For some projects, the most promising sources will be documents published by government statistical agencies. In fact, in the past, students have often restricted themselves to problems that can be answered using data from such sources. But you will also probably want to consider data collected by the various departments of the federal and state governments. The financial statements published by corporations or public utilities can be of use as well. Some researchers actually collect their own data through surveys, or conduct experiments on the Internet. This is a time-consuming and costly way to proceed and requires that you understand how to design and to execute surveys; the sampling procedure as well as the wording of questions can be very important. You should avoid this latter approach, at least until you have gained considerable experience.

2.4. Submitting Your Outline

Your outline should have the following structure: In the first section, provide a brief description of your topic, including some institutional information that will make possible an evaluation of the theoretical model you have chosen, to be presented in a later section. In the next section, present a summary of what other authors have written on your topic. In the third section, present your theoretical model, pointing out the implications of the model for the business problem you are trying to solve. From your theoretical model, then develop an empirical specification within which you can provide a business or economic or behavioral interpretation in the fourth section. In the fifth section, describe the sources from which you intend to draw the data necessary in training the empirical specification. This will involve a complete description of each variable you intend to use from each source as well as a listing of the formulae you intend to use in transforming these variables into measures which are germane to your The final section of the outline should include the list of readings.

Your outline should be printed double-spaced on $8\frac{1}{2}\times 11$-inch paper. The list of readings should be single-spaced with one blank line between references, and generated using the reference management software BibTeX. Leave a one-inch margin around all text. The outline should not exceed twelve pages in length.

3. Implementing Your Outline

Having completed the outline, you will be in a position to execute the research for your capstone project. This will entail inputting your data, analyzing them and interpreting the empirical results as well as writing up these results.

3.1. Inputting the Data

There is no correct way to input data, but several ways exist to organize information so that expanding your dataset with the advent of new data can be done easily. Obviously, building a relational database is your ultimate goal, but start small.

Input data into computer files in an orderly manner. In some cases, for instance, orderly could just mean storing each data series in a separate file. For example, you could use the csv format, which loads into spreadsheets (like Excel) easily, and can be brought into SQLite with a minimum of fuss or bother, too.

As was recommended in ECO 5445, you should develop a convention for naming files. For example, if you have many files in which you store data in csv format, then these files might all have the suffix `.csv`. Provide meaningful prefix names for your files, too. For example, instead of `my.out`, denote the output from estimating a log-linear regression model by `LogLinearRegression.out`. Also, collect your work concerning this project in a folder (subdirectory), within which other folders (subdirectories) also live. For example, the folder could be named `Project` and within it could live the folders `Code`, `Data`, `Docs`, `Figures`, and so forth.

Use the `make` tool (or, at the very least, shell scripts) to automate the process. In the face of frequent revisions, which will inevitably happen as you build your dataset, this should speed-up the process of updates and reduce the number of errors made during that process.

At the risk of belaboring the obvious, maintain daily backups of your files as you proceed, using a shell script to automate the process so it will be easy to do. Then, when disaster strikes, you will have lost at most one day's work.

3.2. Analyzing the Data

This section is deliberately brief because I shall discuss at length specific techniques for analyzing data during the course. Nevertheless, some general guidelines exist. Examine your data in a systematic manner. First, consider the univariate descriptive statistics for each series. Then plot the histograms for these series. If your data are time series, then graph them against time to see if there are any trends or seasonal components. Consider, too, the graphs of the dependent variable versus each of the potential regressors as well as the correlation matrix for all of the data. The purpose of these exercises is to provide you with some feeling for the data.

Next, estimate a simple linear specification of your empirical model including the important variables. Check to see whether the estimated parameters of this specification are sensitive to adding other explanatory variables. Check to see whether the parameters vary across subgroups of the sample. Also check to see whether the assumptions explicit

and implicit in both the analytical and the empirical specification are consistent with the data. One way of doing this is to examine the fitted residuals from the regression in various ways.

Finally, entertain alternative empirical models and examine the predictions of these specifications in light of the predictions of your baseline model. Again, for every specification you try, check to see if the assumptions explicit and implicit in both the analytical and the empirical specification are consistent with the data.

3.3. Writing Up Your Results

Presenting your results clearly and succinctly is as important as the modeling, data gathering, and calculations undertaken. In fact, if your paper is not written clearly and succinctly, then your research will have been for nought. I encourage you to spend considerable effort in writing up your results, thoughtfully. I shall penalize you for mis-spelt words and poor grammar as well as awkward phrases or sentences and poorly-structured arguments. Therefore, consult a dictionary when the spelling of a word is in doubt and a thesaurus when a synonym is required. If you are concerned about your writing style, then the following book is a good place to start:

> Strunk, Wm., Jr. and E. B. White (1972). *The Elements of Style*. Second edition. New York: MacMillan.

But in my book with Golyaev others guides are suggested as well. In all cases, your written work should have the following structure:

- an introduction, where the problem is motivated and the interesting features of the analysis summarized;
- a description of the institutions;
- a development of the economic model;
- a development of the statistical model;
- a description of the empirical specification;
- a discussion of the economic or behavioral interpretations for the model's parameters;
- a description of how the parameters of the empirical model were estimated and a presentation of the results;
- an interpretation of the empirical results in light of the economic and statistical models;
- a discussion of the policy implications of your empirical results;
- a concluding section in which you describe what you learned from the project and summarize your results and their policy implications;
- acknowledgements;
- data appendix;
- any additional appendices, such as those describing numerical or simulation methods used;
- the references.

Readers often find it helpful when the sections of a paper have headings. For example,

> **1 Introduction and Motivation**
> I investigate the effect of ...
>
> **2 Some Institutional Details**
> The industry is best summarized as follows:
>
> **3 Economic Model**
> I begin by employing a simple model of stock market behavior known as the **C**apital **A**sset **P**ricing **M**odel, **CAPM** for short. In this model,
>
> $$\tilde{R}_i - p = \alpha_i + \beta_{mi}(\tilde{R}_m - p) + \tilde{r}_i, \tag{3.1}$$
>
> which means that ...
>
> **4 Statistical Model**
> I map the above economic model into a statistical model ...
>
> **5 Empirical Specification**
> I implement the model developed in sections 3 and 4 as follows:
>
> $$Y_{it} = \alpha + \beta X_{it} + \gamma_i D_{it} + \delta_i D_{it} \cdot X_{it} + U_{it} \quad i = 1, \ldots, k \quad \text{and} \quad t = 1, \ldots, T. \tag{5.1}$$
>
> Here, ...
>
> **6 Empirical Results**
> The parameters of model (5.1) were estimated using ...
>
> And so forth.

3.4. Providing Adequate Recognition to Other Scholars

If you use results that other researchers have derived or ideas they have expressed in their work, then you must acknowledge the contributions of these authors. Such acknowledgements should occur in the text of your paper. In referencing a book, an article, or a chapter in a book adopt the following style:

> Using the reasoning behind the proof presented by Spanos (1987, pp. 329–330), we can show that
>
> $$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top [\mathbf{V}(\hat{\boldsymbol{\theta}})]^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim \chi^2(Q)$$
>
> when the null hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is true.

where the bibliographic citation for this book is

> Spanos, Aris (1987). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Here "pp." means "pages," while in the following citation to a journal article "p." means "page":

> Using the formula derived by Heckman (1979, p. 157), the standard error of the estimate is 0.35.

where the bibliographic reference is

> Heckman, James J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47, 153–161.

In referencing an article in an edited volume, adopt the following convention:

> Thus,
>
> > All the techniques of demand analysis so far discussed share a common approach of attempting to fit demand functions to the observed data and then enquiring as to the compatibility of these fitted functions with utility theory.
>
> as Deaton (1986c, p. 1796) has pointed out.

where the bibliographic reference is

> Deaton, Angus (1986c). "Demand Analysis." In Zvi Griliches and Michael Intriligator (Eds.) *Handbook of Econometrics*, Volume III, pp. 1767–1839. New York: North-Holland.

Here, the "c" after 1986 denotes that there are several articles by Deaton for the year 1986 in the bibliography, this being the third. In referencing information taken from a government publication, adopt the following convention:

> These statistics were compiled by Statistics Canada (1983, p. 10).

In this case the bibliographic reference is

> Statistics Canada (1983). *Logging: Annual Census of Forestry*, catalogue number 25–201.

A reference to a web page was demonstrated above.

Failure to make these acknowledgements is plagiarism, the gravest of academic crimes. At this university, the penalty for either substantial or complete plagiarism is suspension.[2] Moreover, this crime will likely be noted on academic your record and can, therefore, prevent you from being admitted to another academic program. Many employers equate plagiarism with theft (the theft of ideas) and will not hire persons found guilty of this offense.

---

[2] Substantial plagiarism is when no recognition is given to an author for phrases, sentences, or ideas, while complete plagiarism is when an entire document is copied and presented as original work.

3.5. Documenting Your Data Sources

In all scientific endeavors, it is important that other researchers not only be able to replicate the work completed, but that they should also be able to build on it. In empirical work, this means that these researchers must be able to reconstruct the data you used and then to employ them as they see fit. For this to be possible, you must document the construction of your dataset clearly and completely. This will involve providing the following:

(1) the correct bibliographic references, as described in section 2.1;
(2) a complete definition and description of all variables used;
(3) a complete description of the formulae used in manipulating the raw data to obtain the variables you ultimately employed;
(4) a summary table of descriptive statistics for the variables used in your work.

This sort of description should be done in a **Data Appendix** and presented at the end of the paper—after any acknowledgements, but before the references.

4. Finished Product

Your completed paper should be printed on white paper $8\frac{1}{2} \times 11$ inch paper in black ink using a pitch of eleven point or more. The text of the paper and any appendices should be doubled-spaced. Footnotes and citations in the bibliography should be single-spaced, with a space between notes or citations. Footnotes must be numbered consecutively, starting at one, while the references must be in alphabetical order. All pages should have a one-inch margin all around the text. Pages should be numbered using Arabic numerals in the center of the page at the bottom, beginning with the number 1. The sections of the paper should be in the following order:

- text of paper
- acknowledgements
- data appendix
- other appendices (if used)
- references.

You must use the LaTeX macros written in the TeX language to typeset the term paper. You may find the `article` document class the easiest class to work with, but some other classes can work well, too. If you use other classes, then the conditions listed above must still be satisfied. As noted above, you must also use BibTeX to manage the bibliographic references. Any figures used must be created in Python, and then exported to either EPS or PDF formats, and then imported into your LaTeX document. Simple, detailed examples of how to do this are provided in Chapter 11 of the book by Golyaev and myself.

This paper is due on 26 July 2024, at 9:00 a.m.. No extensions will be given: papers submitted after this date and time will be penalized at the rate of twenty percent of the total points remaining per fraction of a day late. This means that after one late day the maximum mark out of 100 you can earn is 80; after two late days the maximum is less than 64; after three late days the maximum is about 51; and so forth.

You must also submit electronically with your paper the source programs and their output as well as the final data files—collected in a `zip` file. Be sure to annotate these documents fully, since marks will be subtracted for things that are unclear.

5. Important Dates

(1) One-page (or less) topic description due: 27 May 2024.
(2) Twelve-page (or less) outline due: 17 June 2024.
(3) Paper due: 26 July 2024.