

SENIOR THESIS HANDBOOK

TABLE OF CONTENTS

I. INTRODUCTION

II. PLANNING FOR THE THESIS TOPIC

III. TIMELINE FOR THE THESIS

IV. DISCUSSION OF STEPS OF RESEARCH PROCESS

1. Developing an effective research question
2. Surveying the literature on the topic
3. Analyzing the issue or problem
4. Collecting relevant data
5. Testing your analysis based on the data
6. Interpreting the results and drawing conclusions
7. Communicating the findings of the research process
8. Oral Presentation

V. SENIOR THESIS WRITTEN ASSIGNMENTS

1. Concept Paper
2. Thesis Proposal
3. Senior Thesis Paper: Drafts and Final Thesis

VI. PLAGIARISM

VII. EVALUATION OF THE THESIS VIII DOING MORE WITH YOUR THESIS

APPENDIX A: EXAMPLE OF THESIS TABLE: VARIABLE DEFINITIONS,

SUMMARY

STATISTICS AND DATA SOURCES

APPENDIX B: STYLE AND DRAFT OF THE THESIS PAPER

APPENDIX C: FREQUENTLY USED STATA & SAS COMMANDS & PROCEDURES

APPENDIX D: THESIS ADVICE FROM ECONOMICS SENIORS

I. INTRODUCTION

This Handbook provides information and advice relevant to undergraduate students in economics for the Senior Thesis. You should read it thoroughly and understand its contents; if you have any questions about it, you can ask in the ECO 494 class or ask your thesis advisor.

II. PLANNING FOR THE THESIS TOPIC

Even before you take the senior thesis preparation course, start to think about the following:

- What have you learned from your economics courses, theoretical, quantitative and applied? Rereading your papers and projects should be helpful.
- Which assignments interested you most? Which readings might serve as a jumping-off point for your thesis topic? Look through the syllabi for courses you have taken.
- What topics in economics news interest you?

Although you take the Senior Thesis near the end of your time as a student, you should start thinking about possible topics in any upper-level course where a paper, or an article or some topic of discussion is especially appealing. By the Fall semester senior year, you will be enrolled in the senior thesis preparation course ECO 494. Your Thesis topic may require specific economic electives. For example, if you are statistically inclined, you should take a course in time-series analysis in addition to the required Econometrics course before using a time-series data set. You are encouraged to choose a topic that relates to courses you have taken; for example, it would be a mistake to try to pursue an international-economics topic if you have not yet taken International Trade. For the mathematically inclined and/or those headed to a PhD program, you may want to take a course in time series as well as the required Econometrics to do a time-series study. For everyone else, the thesis is a chance to “cook” with statistics. We have prepared a set of “recipes” for the kinds of econometric techniques often required by different kinds of datasets and economic problems, for both Stata and SAS (see Appendix B). Once you have specified your model, you will need to look at these. Often, OLS regression is not sufficient, you are likely to face well-known statistical issues that you will have to address through other techniques.

Selecting an appropriate research topic is one of the most important steps in the successful completion of the Thesis. First you need to consider the general area the project will cover: e.g. unemployment in the US; smoking among teenagers in the U.S., etc. Note that a research topic is a step further than the general area (e.g. Labor Economics, International Trade). You have to identify a research question that your thesis will hope to answer, e.g., to what extent does education level affect the probability of becoming unemployed, or does involvement in extracurricular activities reduce the probability of smoking. An effective strategy is to develop a topic you have previously studied for a paper in another course, either a field course and/or Econometrics, or in a MUSE project. It definitely helps to have a topic which interests you and/or relates to an area in which you hope to be employed. Theses are great talking points for job interviews or graduate-school essays. Remember that your project will probably have a quantitative aspect and that is why a paper from ECO 420 or ECO 231 may be a good starting point since you may have already found a dataset for that project. Finding data successfully, downloading it, and cleaning it (more below) is a very important and time-consuming step in your thesis process.

Once you have determined a general research topic, you need to develop a specific research question, which is the topic of Step 1 of the research process, below.

III TIMELINE FOR THESIS

Planning how to organize your time in the semester before you take your Senior Thesis, when you will be taking ECO 494, Senior Thesis Preparation, is essential. In our experience, students have two major problems with thesis time-management. First, they underestimate the time it will take: you should expect it to take two or three times your initial estimate. Second, students have difficulty scheduling stages of the thesis prospectus and thesis itself effectively: which is why we have set intermediate deadlines which must be met (see Timeline below).

There are several parts to the process of carrying out a Senior Thesis: the Concept Paper (1 p.), develops the research question and identifies the literature and relevant data, the thesis prospectus analyzes the literature and develops your hypotheses; data collection and cleaning and model formulation; statistical analysis and a first draft; improving the statistical analysis by additional techniques; revising the model or developing new proxies for your causal model or adding a control variable; re-writing several times to prepare the final paper; and the oral presentations. Each part needs to be completed in a timely fashion to allow you to finish an acceptable project in the time available. At the same time, one size does not fit all. For some students the data is readily available, for some there are many headaches in finding the data. Sometimes the data needs major cleaning, other times everything needed is available online. Some students have major headaches with the statistical aspect of their analyses, others do not. Some are quite comfortable writing up the statistical results, using published articles as a template, others need time and coaching to figure out how to explain their findings. So you should never feel like you are ahead of the game, because you do not know how much work awaits you at the next stage.

This is a major writing project and economists are notoriously not fantastic writers, so as you prepare your thesis prospectus, read Deidre McCloskey's, *Economical Writing* and Strunk and White's *Elements of Style*, two slim volumes that are full of tips on good writing. Doing this reading in advance will help you when it comes to drafting and redrafting your thesis.

Most of you will be doing your research using econometrics. It may be a year or more since you took that class and your skills may be rusty at best. You should review your notes and text on regression for the kind of dependent variable you're using (time-series, cross-section, panel; level, percent, log, dummy variable) before the ECO 495 semester starts.

The first item in the Timeline is preparing a brief (one-page) Concept Paper (see Section VI for instructions). It is due before you write your formal Thesis Prospectus, so at least a month before the end of the Fall semester, and may take the form of an email to your thesis advisor. That is, you need to define your research topic and research question, identify the relevant literature, and verify data availability before you write up your prospectus.

The following timeline is for the thesis-preparation semester, ECO 494:

Week I - IX Meet with ECO 494 coordinator and other faculty
 Read various literatures, try to identify thesis topic
 Meet with economics faculty members to talk about your topic and get some ideas
 Identify your advisor and your basic topic

Week X Submit Concept Paper to your advisor (this can be via email)

Week XIII Submit initial Thesis Prospectus

Week XIV Submit final Thesis Prospectus,* incorporating advisor feedback

*Meeting the deadline for the Thesis Prospectus is essential, since you will not be signed into ECO 495 if you have not completed a satisfactory Thesis Prospectus.

The following timeline for the senior-thesis semester, ECO 495, has been developed to assist you:

- | | |
|--------------------------------|---|
| Week I | Completed Lit Review, Model, basic data ideas |
| Week II | Your complete dataset, correlation matrix, model revisions
Table/plot of stylized facts for variables (plot if time-series) |
| Week IV | Initial econometric runs, extended data, further model modifications |
| Week VI | Economics Field Exam sometime in this vicinity |
| Week VII | Start writing thesis:
Revised Lit review, focused on your final topic
Statistical results
Discussion of econometric results – tie to main hypotheses |
| Week IX | Complete first draft: This includes
Lit review, final draft
table of stylized facts for your dependent & explanatory variables (mean, SD, trend/changes over time if appropriate)
Preliminary econometric results
write-up of econometric results |
| Week X | Thesis draft # 2, including additional statistical work if necessary |
| Week XI | Prepare PowerPoint presentation for Ursinus – as if you're teaching a class |
| Week XII | Ursinus Conference - present your paper, respond to questions |
| Week XIII | Complete revised draft, based on advisor comments and what you figured out from your Ursinus presentation, discussion and discussant comments |
| Week XIV | Thesis "defense" - present to economics faculty & your friends at TCNJ
Celebration of Student Achievement |
| Week XV, first day final exams | Final Thesis due |

Your advisor may modify the timeline. Completing the first draft of the Thesis in a timely manner is essential. See Appendix C for past thesis students' descriptions of their thesis process and timeframe.

IV. DISCUSSION OF STEPS OF THE RESEARCH PROCESS

***provide urls

V. SENIOR THESIS WRITTEN ASSIGNMENTS

All written work should be dropboxed on SOCS or Canvas as well as submitted by email or paper. Organizational details of each are discussed in this section

A. Concept paper

The concept paper should be approximately 1 page long, may take the form of an email to your advisor, and consist of 3 parts:

- Part 1. Description of the research question.

In this part you should address the nature of the economic problem. You need to briefly explain why the question is worthy of study. Spell out the hypothesis that you plan to test in the project and how it will advance our understanding of the problem (i.e. how it creates new knowledge).

- Part 2. Data description

In this part, indicate what data will be used in the analysis, and the sources you have found for that data. Be specific as to which time periods, states, or countries you have found for which you can get the variables you are exploring. This is most important for your dependent variable and key explanatory variable.

- Part 3. Short bibliography

This should consist of the ten "major" papers written on the topic of your research question. If not much work has been done on your exact research question, find parallel research (e.g., for soccer players' salaries, look at studies of other team sports players). At least one paper should be no more than 2 years old. You may also include relevant news articles, if this is a timely topic, but these do not substitute for theoretical or empirical sources.

B. Thesis Proposal

The formal written proposal to be submitted to your advisor should be typed, double spaced, with headings for each of the following sections.

1. Research Question and Motivation for the Research.
2. Initial Literature Review. Show how what others have found either leads to your research question, or does not control for variables that theory or some other literature suggests be explored.
3. A specific statement of what the paper will demonstrate including the major hypotheses to be examined.

4. An Initial Description of Data & Methodology: Indicate the general model you plan to use in your project, making clear why you are including each of the variables in the model. Provide information on the available data you have found to use with the model. List fully the sources of the data, and provide a table of the descriptive statistics. Your proposal will probably be 6 to 25 pages, double-spaced, including the required table describing your data (means, standard deviations).

C. Senior Thesis Paper: drafts & final thesis

The first draft of your thesis should follow as much as possible the required format for your final version. Note that the technical guidelines for paper writing (written below), apply to both the draft and final paper. They are a modified version of the *American Economic Review* guidelines. It is important to use them in your draft, so you can correct any errors before you hand in the final paper. That includes providing your list of references: do not leave that task to the very end!

The final paper should consist of:

1. A title page (thesis title, your name, the date, professor's name)
2. Abstract.
The abstract is a one-paragraph summary of the hypotheses tested in the paper or ideas and applications developed, the theoretical framework and methodology developed to test them and the major findings of the paper. Journal articles used in your project can provide models of appropriate abstracts. It is the last item you complete for the thesis.
3. Table of Contents.
4. Introduction to the Paper.
This opening section should provide: (a) motivation for the paper, offer background and explain why the topic is important; (b) an exact and complete statement of purpose which usually will consist of a set of questions and a concrete hypothesis that the paper will test or answer; (c) a statement of the method the paper will use to test the hypothesis; and (d) an overview of the rest of the paper. You will find discussion of these items under the appropriate headings below. The final form of the introduction is the second last item you complete, just before the abstract.
5. Literature review.
6. Theoretical model development and specification. This ends with spelling out the specific equation you will empirically explore, and the expected signs on your explanatory variables.
7. Econometric results.
8. Interpretation of results.
9. Conclusion and Suggestions for future study
10. Bibliography
Make sure all the work you use in your thesis is listed in the Bibliography. All articles, journals and books must be cited using the format in the *AER Guidelines*. There is no need to cite a website if you have the .pdf of an article from a database that was published in

print. When citing websites, e.g., for a working paper, use the complete URL and please note the date that you accessed the site.

When a site is used as a source of data, provide instructions to actually get to the data you are using. Readers need to be able to find the original data, so that they can replicate results. You may include all of your bibliography for your data in an Appendix table specifying your data sources.

The first draft needs to include items 1 to 7, that is up to and including your first quantitative results. Many of you will be thinking: now I'm done. Alas, that is not true. As everyone who has gone through the process will tell you, now you are beginning. The first results are likely not to be significant, or significant with the wrong signs. Now you have to work out what went wrong and how to change it and you need to start by consulting your faculty advisor.

Revising a draft can make a significant difference to your final grade. That is why four weeks are assigned to it. The document will need three or four drafts: continuous revision is an essential part of the research and research writing process. In the first draft you need to just get down the material: in later revisions you put together the convincing argument to readers of what you have found. You should expect to make major changes in both content and presentation, based on the feedback from your advisor and from your peers. All comments by your advisor on your draft should be read carefully and those comments should be taken into account when making changes. While you need not make every change recommended by your advisor, you should be ready to defend your decision not to accept the recommendations. You can expect your advisor to provide general and specific evaluative comments on the content of your draft but no detailed instructions for revising it: that is up to you. Look carefully for comments on omissions and do substantive work to deal with them.

If there are comments on grammar and clarity, then you must re-write. Remember that most writing goes through numerous re-workings for clarity and improved style. Your writing is supposed to reach professional standards.

Note that examination and return of a working draft does not imply that the paper is acceptable. The paper is never officially judged acceptable or unacceptable until it is in its final and complete form and has been read by both readers and the oral presentation has been made. Helpful comments at Ursinus can show you something to improve the thesis.

Make sure your final thesis meets all the style criteria and is a professionally presented piece of work, submitted by the required deadline.

D. Oral Presentations

After you submit your final paper, you will make two formal oral presentations of your topic to the Ursinus Conference, and then at the Celebration of Student Achievement to the economics faculty and your peers/friends. Each presentation should be about 15 minutes: 10 minutes for presentation and 5 minutes for questions.

E. Thesis Report for next year's seniors

Please write up a summary of your thesis process, how long it took to find your topic and how you finally narrowed it down, how long to write the first lit review (give them dates, it will help

them), rewrite it in the spring, get the data, revise your model, run the regressions, get feedback, finish up. We're trying to help students take 494 more seriously and get them started earlier to help people finish on time. What could someone have told you that would have helped you nail down your topic sooner, or do other parts more easily? Feel free to talk about work responsibilities or access to campus computers or whatever else will give them a sense of reality about what it takes to do the thesis.

VI. PLAGIARISM

[***provide citation/url]

VII. EVALUATION OF THE THESIS

For an expanded view of what economics faculty look for, [***give citation/url]. The basic framework used by the department is provided below.

Senior Thesis Assessment - Written and Oral Communication

Objectives	Low Performance (1 point)	Acceptable (2 points)	Exemplary (3 points)	Earned Points
Written Communication				
1. <i>Research Question</i>	No research question.	Clear statement of the research question and some discussion of the relevance of the question.	Clear statement of the research question and discussion that builds a strong case for the relevance of the research question.	
2. <i>Literature Review</i>	Literature review missing or not relevant to the research question.	Coherent literature review relevant to the research question but some small problems (i.e., not comprehensive or current).	Coherent literature review relevant to the research question that is comprehensive and current.	
3. <i>Support for Conclusions</i>	Unsupported or no conclusions.	Conclusion that highlights key results.	Conclusion that highlights key results and makes key points for the broader significance of the results.	
Oral Communication				
1. <i>Voice Quality and Pace</i>	Demonstrates one or more of the following: mumbling, hard-to-understand English, too slow, too fast, verbal filler.	Can easily understand - appropriate pace and volume. Delivery is mostly clear and natural.	Excellent delivery. Conversational, modulates voice. Projects enthusiasm, interest, and confidence.	

2. Use of Media/Rapport with Audience	Relies heavily on slides or notes. Makes little eye contact. Inappropriate number of slides	Maintains eye contact with audience 90% of the time. Appropriate number of slides. Looks at slides to stay on track.	Perfect eye contact. Slides used effortlessly to enhance speech.	
3. Ability to Answer Questions	Cannot address basic questions.	Can address most questions with correct information.	Answers all questions with relevant, correct information. Speaks confidently.	
				Total

Outside Reader's Scores (1 to 3 scale) on Senior Thesis Assessment

(3 = exceeds expectations, 2 = meets expectations, 1 = falls short)

Written Communication

Score

1. Definition of research question
2. Quality of literature review
3. Support for conclusions

Statistics

1. Variable definitions and relevance
2. Appropriate analysis and statistical diagnosis
3. Explanation of results
4. Support for conclusions

Oral communication

1. Voice quality and pace
2. Use of Media/rapport with audience
3. Ability to answer questions

VIII. DOING MORE WITH YOUR THESIS PROJECT

[***provide citation/Url]

Appendix A

Variable Definitions, Summary Statistics and Data Sources: An Example

Variable	Definition [mean; standard deviation]	Source
<i>LCigSales</i>	Taxable cigarette sales per capita by state (packs). [4.34; 0.37]	Orzechowski and Walker (year)
<i>LCigPrice</i>	Retail price per 20-pack of cigarettes, by state, in 1982-84 dollars based on the Consumer Price Index (CPI) (cents/pack). [5.06; 0.28]	Orzechowski and Walker (year)
<i>Internet</i>	Percentage of state households with internet access. [0.43; 0.22]	Current Population Survey (CPS) and Goolsbee, et al (2010). CPS data available for 1993, 1997, 1998, 2000, 2002, 2003, 2007, 2009. Estimates for other years are interpolated from these data.
<i>LINCpc</i>	Per-capita state disposable income (deflated by the CPI). [9.56; 0.16]	Bureau of Economic Analysis
<i>LBorderP</i>	Minimum retail price in geographically contiguous border states (deflated by the CPI). [4.96; 0.26]	Author's calculations based on above data.
<i>RegShip</i>	State bans the delivery of cigarettes directly to consumer. [0.04; 0.20]	Chriqui, et al (2008)
<i>RegEvade</i>	State has law to deter tax evasion by internet vendors of cigarettes. [0.24; 0.43]	Chriqui, et al (2008)
<i>Mexico</i>	Dummy variable identifying states sharing border with Mexico. [0.08; 0.27]	

<i>Canada</i>	Dummy variable identifying states sharing border with Canada. [0.21; 0.41]	
<i>Producer</i>	Dummy variable identifying main tobacco producing states: Georgia, Kentucky, North Carolina, South Carolina, Tennessee, Virginia. [0.13; 0.33]	
<i>Casino</i>	Dummy variable identifying states housing one or more American Indian casinos. [0.81; 0.40]	Goolsbee, et al (2010)
Notes: The data include annual state level observations from 1994-2009. All monetary variables deflated by the Consumer Price Index (1982-84 = 100). Alaska and Hawaii excluded from the data set (no border states). Prefix L denotes a natural logarithm.		

APPENDIX B

STYLE OF DRAFT AND FINAL THESIS PAPER

Both your *draft* and *final* thesis must meet the following guidelines.

1. Use only standard-size paper (8.5 x 11 inches). Use a 12-point Times New Roman font and maintain a 1-inch side, top, and bottom margin. Number all your pages, excluding the title page .

2. **DOUBLE-SPACE** the entire text.

3. **TITLE PAGE** should be on a separate page with thesis name, author, date, thesis advisor, abstract, and any acknowledgements.

4. **TABLE OF CONTENTS** is required and should note headings and sub-headings by page number.

6. **SECTION HEADINGS AND SUBHEADINGS** should be clearly detailed in your paper. These headings should have a physical layout that helps the reader comprehend the structure of the paper. Make the headings informative. Section headings should be centered and given Roman numerals (I, II, etc.); subsections should be lettered A., B., etc. and are left-oriented.

7. **ENDNOTES** must be double-spaced. The footnotes should be numbered consecutively (i.e., 1, 2, 3, etc.).

8. **REFERENCE TO INDIVIDUALS IN THE TEXT** should include the first name, middle initial, and last name in the first instance. Subsequent references should give last name only. Do not refer to individuals as Mister, Doctor, Professor, etc.

9. **REFERENCE TO ORGANIZATIONS OR GOVERNMENTAL AGENCIES IN THE TEXT** should give the name in full, followed by the abbreviation in parentheses -- subsequent references should give abbreviation only; for example: Social Science Research Council (SSRC) [first occurrence], SSRC [subsequently].

10. **REFERENCE TO ARTICLES AND BOOKS IN THE TEXT:** Give full name (first name, middle initial, and last name) of author(s) and year of publication in the first citation, with page number(s) where appropriate. Do not include the title of articles in the text, that is for your bibliography. When more than one work by the same author is cited, give the last name of author and year of publication in parentheses for each subsequent citation. When listing a string of references within the text, arrange first in **chronological** order, then alphabetically within years. If there are three or more authors, refer to the first author, followed by et al. and the year. If there is more than one publication referred to in the same year by the author(s), use the year and a, b, etc. (example: 1997a, b). References to authors in the text must exactly match those in the Reference section.

11. **REFERENCE TO INFORMATION ON THE WEB:** When citing Internet sources as a general rule include information on the (a) author(s) or publishing agency, (b) title of work, (c) print publication information, (d) title of online site, project, journal or database, underlined, (e) the date when the work was posted or updated electronically, (f) date when you accessed the site, and (g) electronic address (URL). For more information on citing Internet and other electronic sources refer to the MLA Web site at <http://www.mla.org>.

12. MATHEMATICAL EQUATIONS should be typed on separate lines and numbered consecutively at the left margin, using Arabic numbers in parentheses.

13. QUOTATIONS must correspond exactly with the original in wording, spelling, and punctuation. Page numbers must be given. Changes must be indicated: use brackets to identify insertions; use ellipsis dots (...) to show omissions. Also indicate where emphasis **has been added** (emphasis added). Only lengthy quotations (more than 4 lines) should be separated from the text; such quotations must be double-spaced and indented at the left margin.

14. TABLES must be on separate pages at the end of the thesis – not incorporated within the text – and should be numbered consecutively with Arabic numbers. In the text, after the paragraph where you first mention table 1, insert the following:

(table 1 goes about here)

Each table must have a title and should be no more than 10 columns wide. Use Panel A and Panel B to denote sections of a table. Do not abbreviate in column headings, etc. Spell out "percent"; do not use the percent sign. Place a zero in front of the decimal point in all decimal fractions (i.e., 0.357, not .357).

Table footnotes are also to be single-spaced. For footnotes pertaining to specific table entries, footnote keys should be lowercase letters (a, b, c, etc.); these footnotes should follow the more general table Note(s) or Source(s). Use asterisk (*) footnotes for the following: *Significantly different from 0 at the 1-percent level (or ** or *** or such other symbols as ⁺ or [^] for other significance levels). Full citations of the sources are to be included in the References.

15. Each GRAPH or FIGURE should fill a page. Graphs are included after the body of the paper. They should be introduced in the text following the paragraph first referring to them.

(figure 1 goes about here)

Also label all axes, curves, and intersections in the graphs, number them consecutively, and provide full footnotes on the graph page wherever relevant. All graphs should have a clear number, title and source attribution.

16. REFERENCE SECTION must be double-spaced, beginning on a new page following the text, giving full information. Use full names of authors or editors (last names first), using initials only if that is the usage of the particular author/editor. List all author/editors up to/including 10 names. Authors of articles and books and material without specific authors or editors, such as government documents, bulletins, or newspapers, are to be listed alphabetically by publishing agency.

A) **Books:** List Author or Editor. Title. Place of publication: Publisher, year.

B) **Articles:** List Author. "Title of Article." Journal, month and year of issue, volume (and issue number) in Arabic numerals, inclusive of page numbers. Specify if volume is part of title (volume 2) or not (Vol. 2).

C) **Unpublished Papers:** List Author. "Title." Working paper or discussion paper (including number if any), institutional affiliation, date.

D) **Chapters in Edited Volumes:** List Author. "Title," Editor, Volume Title. Place

of publication: Publisher, year, inclusive page numbers. Specify if volume is part of title (volume 2) or not (Vol. 2).

17. DATA SOURCES must be given with full information and listed in the References or in a separate Data Appendix.

18. OTHER STYLE POINTS: (1) In the acknowledgement footnote on the title page, feel free to acknowledge your thesis advisor and anyone else you feel has helped you with your paper. (2) Do not use the % sign in the text; always spell out the word percent; (3) Apostrophes are used for decades (i.e., 1990's), not generally for plurals (i.e., HMOs); (4) Hyphenate compound adjectives when they come before a noun, not after (i.e., a well-known author; an author well known).

19. LENGTH of the paper may be approximately 15-45 pages.

Formatted: Tab stops: 2", Left

FREQUENTLY USED STATA & SAS COMMANDS & PROCEDURES

General:

Formatted: Font: Not Bold

.PWCorr y x z gives pairwise correlations for the 3 variables y, x and z.

.Regress y x z regresses variable y on variables x and z

.Stepwise, pe(.20): regress y x z

Runs a stepwise regression for y on variables x and z, adding in variables significant at least at the 20% level

Time-series Data:

Formatted: Underline

.Tsset year To set your time-series variablename (assuming you called your years "year")

Then after any regression you want to check for autocorrelation, immediately type

.Estat dwatson

And it will give you Durbin-Watson stats. You'll have to look up the relevant range for DW significance, given your # regressors and sample size.

To run an equation corrected for Durbin-Watson issues, type

.Prais y x z To regress y on variables x and z, correcting for autocorrelation

Time-series Data and Stationarity

.Dfuller x

- gives Dickey-Fuller stat for variable x. Checks for stationarity of trend variables.

Notice, more strongly negative MacKinnon statistics mean stationarity is less likely.

If your variables are non-stationary, you'll have to calculate "changes in" for all your variables, and run Dickey-Fuller tests on these "change in" variables.

Cross-section Data:

Formatted: Underline

Heteroskedasticity

Formatted: Font: Bold, Underline

Immediately after you've run a regression, type

.Estat hettest

This will let you know the probability of no heteroskedasticity – if your probability is very low, you need to estimate robust coefficients. Type

.Regress y x z, robust

Adding the **robust** command in a regression corrects for heteroskedasticity.

Panel Data

Formatted: Underline

First, make sure you have told Stata what your cross-section & time-series variable names are, e.g., but typing

.Tsset StateID Year [or .Xtset StateID Year]

Or

.Tsset CountryID Year [or .Xtset CountryID Year]

You need to include a StateID variable which translates state names or abbreviations into a numerical equivalent; for instance,

<http://www.epa.gov/enviro/html/codes/state.html>

provides the FIPS codes for each state, which is a number.

Field Code Changed

If you have panel data, you want to control for fixed effects. Suppose you have state data. There will be cross-section idiosyncracies, such as some states always tending to have higher or lower values of your dependent variable than others because of unspecified state-characteristic variables.

Hausman Test for Random Effects vs. Fixed Effects

Formatted: Font: Bold

If you have panel data, you may want to control for fixed effects. Suppose you have state data. There will be cross-section idiosyncracies, such as some states always tending to have higher or lower values of your dependent variable than others because of unspecified state-characteristic variables.

Then ~~run~~ run

.Xtreg y x z, i(StateID) fe

. estimates store fixed

. Xtreg y x z, i(StateID) re

Then test the appropriateness of the random-effects estimator (xtreg, re) by running

. hausman fixed ., sigmamore

This runs a regression while allowing you to control for fixed effects (via fe command), shows you whether you need to use fixed effects.

Testing panel data for heteroskedasticity

The two commands

_____ . xtgls ... , igls panels(heteroskedastic)
_____ . estimates store hetero

fits the model with panel-level heteroskedasticity and saves the likelihood.

We can fit the model without heteroskedasticity by typing

_____ . xtgls ...

Now there is one trick. Normally, lrtest infers the number of constraints when we fit nested models by looking at the number of parameters estimated. For xtgls, however, the panel-level variances are estimated as nuisance parameters and their count is NOT included in the parameters estimated. So, we will need to tell lrtest how many constraints we have implied. The number of panels/groups is stored in e(N_g) and, in the second model, we are constraining all of these to be single value, so our number of constraints can be computed and stored in a local macro by typing

Formatted: Tab stops: Not at 0.64" + 1.27" + 1.91" + 2.54" + 3.18" + 3.82" + 4.45" + 5.09" + 5.73" + 6.36" + 7" + 7.63" + 8.27" + 8.91" + 9.54" + 10.18"

Field Code Changed

_____ . local df = e(N_g) - 1

The test is then obtained by typing

_____ . lrtest hetero ., df(`df')

Testing panel data for autocorrelation

There is a user-written program, called xtserial, written by David Drukker to perform this test in Stata. To install this user-written program, type

_____ . findit xtserial
_____ . net sj 3-2 st0039
_____ . net install st0039

To use xtserial, you simply specify the dependent and independent variables:

_____ . xtserial depvar indepvars

A significant test statistic indicates the presence of serial correlation.

Correcting panel data for autocorrelation, controlling for fixed effects:

. xtregar depvar indepvar1 indepvar2, fe

Correcting panel data for both heteroskedasticity and autocorrelation

. xtgls depvar indepvar1 indepvar2, panels(hetero) corr(ar1)

Resources:

Time series Data and Stationarity

Dfuller x

gives Diekey Fuller stat for variable x. Checks for stationarity of trend variables.
Notice, more strongly negative MacKinnon statistics mean stationarity is less likely.

If your variables are non stationary, you'll have to calculate "changes in" for all your variables, and run Diekey Fuller tests on these "change in" variables.

Formatted: Underline

Special Dependent Variables: Proportions; Dummy Variables

Proportions as Dependent Variable

Natural Logs

If your dependent variable is a proportion, whether 0 to 1 or 1% to 100%, then it is truncated - it can't be more than 1 (100%) or less than 0 (1%). So for consistent estimation, you need to log your dependent variable, which will spread out the range of the variation. If any of your regressors (explanatory variables) are also proportions, log them as well so they are being handled in a symmetric fashion. You can then run OLS regressions on this logged data.

Logit Transformation

If you take the log of (your proportion/(1-your proportion)), this transformed variable has the merit of looking much more like a straight line than a simple proportion could – the latter is subject to a binomial distribution. So many researchers use a logit transformation on proportions, on both dependent and explanatory variables which are proportions.

What to do if any of your proportions actually equal 0

The solutions based on logs or logit transformation will not work if any of your answers are actually 0, because the log of zero is undefined. Then you will have to use a generalized linear model or GLM estimation to prevent biased coefficient-estimates on your raw data. This is a form of maximum-likelihood estimation:

.GLM y x z

Gives you maximum-likelihood estimates for the effects on y of x and z.

Dummy Variable as Dependent Variable

If your dependent variable is either 0 or 1 (e.g., you are or are not a juvenile diabetic), you have to perform a **probit** regression:

.Probit y x z

Will calculate regression coefficients for your kind of dependent variable.

~~Panel Data~~

~~First, make sure you have told Stata what your cross section & time series variable names are, e.g., but typing~~

~~**.Tsset** StateID Year [or **.Xtset** StateID Year]~~

~~**Q#**~~

~~**.Tsset** CountryID Year [or **.Xtset** CountryID Year]~~

~~You need to include a StateID variable which translates state names or abbreviations into a numerical equivalent; for instance,~~

~~<http://www.epa.gov/enviro/html/codes/state.html>~~

~~provides the FIPS codes for each state, which is a number.~~

Field Code Changed

If you have panel data, you want to control for fixed effects. Suppose you have state data. There will be cross-section idiosyncracies, such as some states always tending to have higher or lower values of your dependent variable than others because of unspecified state characteristic variables. Then run

```
xtreg y x z, i(StateID) fe
```

This runs a regression while allowing you to control for fixed effects (via fe command).

~~Testing panel data for heteroskedasticity:~~

```
xtgls var, igls panels(heteroskedastic)  
estimates store hetero
```

fits the model with panel level heteroskedasticity and saves the likelihood.

We can fit the model without heteroskedasticity by typing

```
xtgls var
```

Now there is one trick. Normally, `lrtest` infers the number of constraints when we fit nested models by looking at the number of parameters estimated. For `xtgls`, however, the panel level variances are estimated as nuisance parameters and their count is NOT included in the parameters estimated. So, we will need to tell `lrtest` how many constraints we have implied. The number of panels/groups is stored in `e(N_g)` and, in the second model, we are constraining all of these to be single value, so our number of constraints can be computed and stored in a local macro by typing

```
local df = e(N_g) - 1
```

The test is then obtained by typing

```
lrtest hetero var, df(`df')
```

~~Testing panel data for autocorrelation~~

There is a user-written program, called `xtserial`, written by David Drukker to perform this test in Stata. To install this user-written program, type

```
findit xtserial  
net sj 3 2 st0039  
net install st0039
```

To use `xtserial`, you simply specify the dependent and independent variables:

```
xtserial depvar indepvars
```

A significant test statistic indicates the presence of serial correlation.

~~Correcting panel data for autocorrelation, controlling for fixed effects:~~

```
xtregar depvar indepvar1 indepvar2, fe
```

~~Correcting panel data for both heteroskedasticity and autocorrelation~~

```
xtgls depvar indepvar1 indepvar2, panels(hetero) corr(ar1)
```

Field Code Changed

~~Resources~~

RESOURCES;

Stata’s “Help” feature is very useful. Click “Help” then “Search.” Type in your topic. Choose among the options whichever sounds closest to your topic. At the end of each explanation, Stata typically provides Examples – these are very useful templates for modeling your adoption of that technique.

Princeton University has great Stata web-based help information. Go to http://dss.princeton.edu/online_help/stats_packages/stata/ and search on your topic. Or a Google search on “Stata Princeton your topic” will usually take you there.

Formatted: Indent: Left: 0", Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Font: Bold

TYPICAL SAS COMMANDS:

General Regression: with White test

```
data copper;
input YEAR PRICE Q;
datalines;
~data~
;
```

White Test:

```
data mileage;
input HP MPG MPGF VOL WT SP one;
datalines;
;
proc reg;
model mpg = hp VOL WT SP/r p;
output out=out1 r=r1 p=y1;
run;
data out1;
set out1;
r1sq=r1**2;
ysq=y1*y1;
lr1sq=log(r1sq);
proc reg;
model r1sq=y1 ysq;
run;
```


Testing Restriction Statements:

```
data product;
input YEAR Y X1 X2;
$ define Y = Real Gross Product, Millions of NT
X1 = Labor Days, Millions of Days
X2 = Real Capital Input, Millions of NT $
datalines;
;
proc reg;
model y=x1 x2/r ;
test x1+x2=1;
output out=out1 r=resid p=yhat;
run;
data test;
set out1;
ly=log(y);
lx1=log(x1);
lx2=log(x2);
proc reg;
model ly=lx1 lx2/r;
test lx1+lx2=1;
```

```
run;
proc reg;
model ly=lx1 lx2;
restrict lx1+lx2=1;
run;
proc reg;
model ly=lx1 lx2 / noint;
run;
proc reg;
model ly=lx1 lx2;
restrict intercept=0;
run;
proc reg;
model ly=lx1 lx2;
test intercept=0;
run;
```

WLS For an unknown functional form (From Class):

```
Data;
infile 'a:\SAVING.raw';
input sav inc size educ age black cons;
cards;
;
proc reg;
```

```

model sav =inc educ age / r;
output out=out1 r=resid;
data out1;
set out1;
r1=resid**2;
proc reg;
model r1=inc educ age / r p;
output out=out2 p=yhat;
data out2;
set out2;
g=log(yhat);
sav1 = sav/g;
inc1=inc/g;
age1=age/g;
educ1=educ/g;
proc reg;
model sav1=inc1 age1 educ1;
run;

```

Correcting for heteroskedasticity with WLS:

Data grade;

input obs gpa tuce psi grade letter\$; (\$ is for when data is not a number)

```

y=grade;
x1=gpa;
x2=tuce;
x3=psi;
datalines;
;
proc reg;
model y=x1 x2 x3/r p;
output out=out1 r=r1 p=p1;
run;
data out1;
set out1;
r1sq=r1**2;
p1sq=p1**2;
if p1 le 0 then p1=.01; (THIS LINE FIXES FOR neg Y HAT VALUES WHEN TAKING THE
SQUARE ROOT FOR
WLS)
p2=1-p1;
psq=(p1*p2)**.5;
invp=1/psq;
yinv=y*invp;
x1inv=x1*invp;
x2inv=x2*invp;
x3inv=x3*invp;
proc reg;
model r1sq=p1 p1sq;
run;

```

```
proc reg;
model yinv=invp x1inv x2inv x3inv / noint;
run;
```

DURBIN D TEST

```
data copper;
input YEAR C G I L H A ;
c1=lag(c);
n=_n_;
datalines;
;
proc reg;
model c=g i h l a / dw;
run;
```

DURBIN H TEST

```
data copper;
input YEAR C G I L H A ;
c1=lag(c);
n=_n_;
datalines;
;
```

```
proc reg;
model c=g i h l a c1 / dw;
run;
```

BREUSCH-GODFREY TEST

```
data copper;
input YEAR C G I L H A ;
c1=lag(c);
n=_n_;
datalines;
;
```

```
proc reg;
model c=g i h l a / dw r;
output out=out1 r=r;
run;
```

```
data out1;
set out1;
r1=lag(r);
r2=lag(r1);
r3=lag(r2);
r4=lag(r3);
```

```
proc reg;
model r = g i h l a r1 r2 r3 r4;
```

```
run;
```

```
proc reg;
```

```
model r = g i h l a r1;
```

```
run;
```

Correcting for autocorrelation:

```
proc autoreg;
```

```
model c= g i h l a / nlag=1;
```

```
run;
```

LOGIT MODEL

```
proc logistic descending;
```

```
model y=x1 x2 x3;
```

```
run;
```

(AT END OF DATA)

PROBIT MODEL (for $P(y=0 \text{ given } x)$)

```
proc probit;
```

```
class y;
```

```
model y=x1 x2 x3;
```

```
run;
```

(to get correct values, for $y=1$, just switch the signs on the beta coefficients)

PROBIT MODEL (for $P(y=1 \text{ given } x)$)

```
proc logistic descending;  
model y=x1 x2 x3 /link=probit;  
run;
```

PANEL DATA (POOLED TIME SERIES AND CROSS-SECTIONAL)

```
data Investment;  
input comp$ year y x1 x2;  
label y ='investment expenditure'  
X1 = 'value of firm'  
X2 = 'capital (stock of plant and equipment)';  
cards;  
;  
/* GENERAL MODEL FOR POOLED DATA OF ALL INDIVIDUALS ACROSS TIME*/  
proc reg data=investment;  
model y=x1 x2 /dw;  
run;  
/*INDIVIDUAL COMPANY MODELS*/  
data GE;  
set investment;  
if comp='_GE';  
proc reg;  
model y=x1 x2/dw;  
run;  
data GM;
```



```
set investment;
if comp='_GM';
proc reg;
model y=x1 x2/dw;
run;
data US;
set investment;
if comp='_US';
proc reg;
model y=x1 x2/dw;
run;
data WEST;
set investment;
if comp='_WEST';
proc reg;
model y=x1 x2/dw;
run;
/*setting up dummy variables for when INTERCEPTS VARY ACROSS INDIVIDUALS
ONLY*/
data investment;
set investment;
if comp='_GM' then d1=1;else d1=0;
if comp='_GE' then d2=1; else d2=0;
```

```

if comp='_US' then d3=1; else d3=0;

proc print;

var y x1 x2 d1 d2 d3;

run;

proc reg;

model y=x1 x2 d1 d2 d3;

run;

/*SORTS ALL COMPANIES BY COMPANY (cross section) AND TIME (time series)*/

proc sort;

by comp year;

run;

proc panel;

model y=x1 x2/fixone;

id comp year;

run;

/*INTERCEPTS VARY ACROSS TIME AND ACROSS INDIVIDUALS*/

proc panel;

model y=x1 x2/fixtwo;

id comp year;

run;

```

Fixed and Random Effect Models

```

data comp;

input Country$ Year COMP UN;

Datalines;

;

```

```
proc sort;
by country year;
run;
proc panel;
model comp=un /fixone; (FIXED EFFECT MODEL)
model comp = un / ranone; (RANDOM EFFECT MODEL)
id country year;
run;
```

Sample Complete Program:

```
data stocks;
input RR GROWTH INFLATION;
RR1 = lag(RR);
n=_n_;
datalines;
;
proc reg;
title 'RR on Growth and Inflation';
model RR=GROWTH INFLATION/r;
output out=out1 r=r1;
```

```

run;

data out1; /*Lagged values for RR on Growth and Inflation*/

set out1;

lr1=lag(r1);

lr2=lag(lr1);

lr3=lag(lr2);

lr4=lag(lr3);

proc reg;

title 'RR on Inflation';

model RR=INFLATION/r;

output out=out2 r=sr;

run;

data out2; /*Lagged values for RR on Growth and Inflation*/

set out2;

slr1=lag(sr);

slr2=lag(slr1);

slr3=lag(slr2);

slr4=lag(slr3);

proc reg;

title 'B-G Test for RR on Inflation';

model r1 = INFLATION slr1 slr2 slr3 slr4 ;

run;

proc reg;

title 'B-G Test for RR on Inflation and Growth';

model r1=INFLATION GROWTH lr1 lr2 lr3 lr4;

```

```
run;

proc reg;
title 'D-W Test';
model RR=GROWTH INFLATION RR1 /dw;
run;

proc reg;
title 'D-W Test';
model RR=INFLATION RR1 /dw;
run;
```

You can contact Dr. Letcher and or Dr. Samanta about SAS commands. You can also check SAS documentation online.

Appendix D

Thesis Advice from Economics Seniors

Economics Thesis Student Reports & Advice, Fall 2011-Spring 2012

Jaffer Ahmed Thesis Advice

Preparing and writing the thesis is a long and time-consuming process that requires time management in order to complete it on time. Working on the thesis along with the regular course load is tough so starting to prepare over the summer before taking ECO494 will help greatly in putting you ahead of the schedule. Picking out a topic over the summer will save a lot of time during the semester. After picking a topic, you must make sure its concise enough to write a thesis on. You have to also make sure that there is data for your topic that is available for access. You have to also make sure there is enough data, which can vary depending on your topic. If you can pick out a topic and find a reasonable dataset to start off with during the summer before the fall semester starts, it will make life much easier as you wouldn't have to spend so much time during the semester searching for a good topic that interests you.

For my thesis, finding the topic and data source was relatively easy. However my dataset was not given in a complete set, ready for analysis. The data needed to be taken for 150 observations separately, and combined into one dataset. Finding, combining, and cleaning up the data took the longest time for me. Since this was also during a very busy semester, it took me over half the semester just to organize and clean up the data. One thing you should do is make sure you know how to manage your time well with the semester so you can also spend time working on your thesis. The toughest part is to make sure your dataset is large enough and has no problems. The toughest part of the whole thesis process in my case was not writing the paper itself, but working on the dataset for it. Therefore, one should put aside enough time, and prepare to do a lot of time-consuming work in preparing the data. If your data is good, then analyzing it and writing your paper will be easy.

I didn't start with any one topic; I had a list of about twenty, most of which were thoughts I had while reading books/newspapers. Through some combination of receiving feedback during the thesis prep class, reading academic studies on the topics, and asking myself questions about how I'd test any of them I eliminated everything on my list.

I then tried to search for ideas elsewhere. I looked through the academic journals the library subscribes to. This was a mistake. Most of these articles say nothing exciting (even in the big name journals) and with most pushing 50+ pages, they take *a lot* of time to read through. Also, unless you're comfortable with high level math, the complexity is sometimes too much.

Better places to look are newspapers and magazines. Hal Varian, a famous microeconomist, says that he got most of his fresh ideas from reading newspapers (including sections like the classifieds) and trying to make sense of trends he noticed. You shouldn't just gravitate to the business/economics section of a newspaper/magazine stand. Reading Paul Krugman's column will make you smarter (it will), and may introduce you to some new interesting topics, but most of the stuff that he – and every other economics writer – talks about is established. It will be hard to think of anything fresh to say about that. [Hal Varian has a paper on how to build economic models that's good ([click here](#)). Big-wig Peter Diamond also has something about finding research topics ([here](#)) but his is less helpful. There are probably a bunch of others out there if you just search for them.]

I didn't finalize my topic until Thanksgiving break, and for that I turned to the best source: professors! You should have a pretty good idea of what your professors' research interests are. Talk to them about what yours are and ask them if they have a suggestion.

My topic was the spillover effects of fiscal policy from one country to others. The paper used Vector Autoregression (VAR) analysis, which is a special type of regression for using multiple time series and trying to account for their interdependencies. The countries I considered were the US, Mexico, Canada, Germany, and Australia (more on why I chose these countries later). Basically, I used VARs to see things like how output in Mexico would respond to an increase in government spending in the US.

I didn't know anything about what VARs were when I decided on the topic, so I needed to learn that. The TCNJ library is really lacking in economics textbooks. Most of the textbooks that are in there have the, "old smell" when you open them. This is especially true of books on econometrics, where a lot has changed recently. Still, your professor may have texts on what you need (mine did) and also, there is *a lot* on the internet, especially on college websites. Anyhow, I spent some time figuring out what VARs were, and then did an initial lit review before winter break. I say initial because as I went on I had to keep reading more and more to make sense of what I was doing. When I was doing the lit review it helped me to check the bibliographies of what I was reading. A lot of the articles I read would cite the same few "seminal" papers on the topic. These papers are usually broader, and this can help since some of what you'll read is so specific that it will be hard to see how to apply it to your work.

Now on to why I chose the US, Mexico, Canada, Germany, and Australia... Originally I wanted to choose countries that had close economic ties, as this would be more likely to show whether spillovers exist (if they don't exist for closely tied nations, why would they for anyone else?). As an example, for Mexico the US and Canada are large trading partners. Therefore, you'd expect that if expansionary fiscal policy was pursued in those countries, it would have some positive effect on output in Mexico. I was told that my study would have more robust results if I had a larger data set, i.e. included more countries. Finding data for other countries was a pain. The data I was looking at had to be inflation adjusted and quarterly, which is surprisingly difficult to find, even for developed countries. And this is why Germany

and Australia were included; they had the data (also, they're pretty significant economies, so not terrible to include). Something I wish I thought of: if you're having trouble finding data but know of another study that used the same data you're looking for (or something similar), try contacting the person who wrote the study. Most academics are congenial about helping one another and their contact information is easy to find.

In terms of dates when I finished certain tasks...ugh...is, "later than I should have", a date? At the time, I was balancing the thesis with four other classes, and a senior capstone in another major. Too much. Also, deciding on a topic so late, and then having to learn so much new material for it was rough. As an example of where I was and when, I ran regressions and found my results in the days before the Ursinus presentation. I started to make my powerpoint for the presentation *an hour* before we left TCNJ for Ursinus (seriously). Thankfully there were only six people in the room when I presented, and I was able to use confusing enough language that they weren't sure what to grill me on during questions and answers. [Don't be afraid about presenting at Ursinus. It's a chance to have fun and share what you worked on. No one in the audience grills anyone. Most of the other kids there are scared (or sleeping). No one's thesis grade is made or destroyed there, either.]

About my results... they were pretty inconclusive. For a few countries, there was evidence of fiscal policy having spillover effects (Mexico, Australia), while for the others there was either no evidence, or not enough to conclude anything. This is a little disappointing, as you're hoping that your results will be a clean confirmation of your hypothesis, but there is no shame in inconclusive results. A lot of things in economics are unclear. Also, it's better to have inconclusive results than to manipulate your data in such a way as to get results, and then cover up all the uncertain results.

Between the Ursinus presentation and the Celebration of Student Achievement I wrote my entire thesis. This is a lot of work, but I'm fast writer, and once you've done a presentation on something it's easy to write about it (although it's certainly better to write about something and then do the powerpoint about it).

The Celebration of Student Achievement was a little bit of a disappointment. I was the first to present (actually there was another presentation at the same time in another room). There were a total of *four people* there to see me present: my advisor, two other professors (one of whom fell asleep during the presentation), and the student who presented right after me. This was a shame, because I was sharp for this presentation and proud of my work. I think the economics department should revise the way they do this to encourage a better turnout. For example, the math department does their capstone talks in the lunch periods (11:20-12:30) during the last few weeks of the semester, with 2-3 students presenting each day. Faculty are encouraged to attend, but so are students – including underclassmen. The talks get 20-30 people, lively Q&A, and actually make you feel like you're holding a lecture.

Overall, I really liked the thesis project. I learned more than I did in ~~most~~ any of my classes, and really liked what I was working on. Also, at the end, the feeling of accomplishment is super. I wish I started earlier; it sucks spending "senior nights" in the library trying to figure out why SAS won't work. Otherwise the thesis was super. I know the econ department allows students to do "independent study" in economics as an elective (although I don't know anyone who did it), and the thesis made me wish that I had done one of those earlier in college.

Pick something you're interested in – even if it's difficult, work hard – it's the last thing you'll do so make it good, and **have fun!** ☺

Alex Durante

Personal Timeline

I began searching for a thesis topic during the fall semester of 2011, as part of the Thesis Preparation course. I encountered multiple difficulties during this process, and switched topics several times before finally settling on one early in the spring semester of 2012. Some of the problems I faced were that I either selected topics that had already been investigated multiple times, or ones that would require data that was not available because they had never been investigated. I would advise students to find a topic that has a substantial body of published research and accessible data, but provides an opportunity to approach the topic from a fresh and interesting angle. Alternatively, students might also want to search for studies that would benefit from improvements in the research design and data construction, even if the angle remains the same.

I wrote the introduction and literature review of the thesis during the winter break of the fall semester. I found this to be one of the least stressful aspects of writing the paper, since this merely requires the ability to summarize the relevant findings of other research. I spent the spring semester compiling data and running regressions, which proved to be quite time-consuming. While some of this data was easily accessible, it often required formatting and reorganization to ensure its compatibility with STATA and my statistical model. Additionally, it was not always clear what STATA commands would yield the desired results. Though I did notice that only a rudimentary understanding of regression analysis is needed to complete the thesis, I would have preferred if the econometrics class I had taken the previous year had been focused more on the application of the concepts that we learned. For instance, it would have been helpful to learn how to control for autocorrelation in a fixed-effects model using STATA.

Though I am somewhat embarrassed to admit this, I did not complete the thesis until the spring of 2013. I strongly encourage to students to plan wisely so that they finish their thesis upon graduation. I would have much rather completed the thesis sooner than have this looming obligation. Writing the data analysis and conclusion sections of the paper are not terribly difficult once you have learned how to use STATA. Overall, I found managing my time more challenging than actually writing the paper.

Nick Falcone
Thesis Report

Timeline and Details

Topic and Literature Review

My ultimate topic was inspired by an HBO documentary I watched the summer before senior year. I ended up having two ideas, but choosing one by the middle-end of the fall semester. I started the literature review around then and finished it two days after the semester ended. I would advise against that.

Data

I began looking for data toward the first or second week of the spring semester. I got lucky and did not have much trouble finding what I needed but I had enough time to accommodate a longer, more tedious search (which I hear is more typical). I had a complete dataset at some point in late February or early March and started running regressions about one week later. I did a logit transformation of my dependent variable because it was a proportion. Transformations, to me, are not as intimidating as they seem – just some time and a few maneuvers in Excel. Be able to explain why you used the transformation you used (if one was necessary).

I had some results to present at Ursinus. Make sure you do as well, because presenting a paper without at least initial results is sort of embarrassing. It's also helpful in terms of time. The two weeks after the conference will be very hectic if they contain all of your data analysis. My presentation at the Celebration of Student achievement contained my final results. I still needed to interpret them in the paper afterward, but the Stata work was complete.

Final Thoughts

“Start early” is a big cliché, but it's true. Give yourself the time to do things the right way. In terms of *how* to do things the right way, I would suggest thinking deeply about what you want your model to say. Think about the relationship you're testing and what measures, structure, controls, etc. make the most sense. Make it as complete as possible (within reason) and try different things. If I could have done one thing differently it would have been to use more controls and not use two of the variables that ended up in my final model.

The thesis is a big undertaking but also a great opportunity to make a contribution and communicate your point of view. If you think of it more in these terms and less as an obligation, everything should fall into place.

Scott Francis
Thesis Process Summary

I started the Thesis process in the fall of 2011 with my Thesis Preparation Class. After attending the first couple of classes I had an idea of what I wanted the subject of my thesis to be. I started to skip the Thesis Prep classes, because I thought I was done my thesis prep since I had decided on a topic. I put my thesis on the back burner and concentrated on my other classes and work thinking I could just write my prospectus really quickly before the end of the semester. This was a big mistake! I put it off until the end of the semester and then realized that finding sources that apply to my topic wasn't very easy. I was then dealing with finals, work, and trying to scramble and find sources for my literary review. This led to me throwing together a bad literary review with some sources that weren't very strong, and I had to rewrite my prospectus to better reflect my topic and data. The semester was over so I had to spend my Christmas break doing all of the work I should have been doing during the fall semester. This is not something I recommend doing and it just weighed down all me all break thinking about how I was behind and not finished. "What if I don't graduate?" is constantly running through your head and it's not a fun thing to think about.

I then started my Thesis class in the spring of 2012 after working all break to catch up with my classmates. I decided to get proactive this time and looked up a bunch of data that I thought would work for my topic. Thinking I was ahead of the game and that I only had to run some easy regressions now I put my thesis on the back burner and concentrated on my classes and was needed at work more. The end of the semester came around and I was just getting started running my regressions. I really looked hard at my data and realized that my data didn't match up. I was in a horrible situation. I ended up having to find all new data and I was supposed to be done my thesis already. Everyone is talking about their jobs and what they are going to do after graduation, and I was thinking I still don't have all my data for my thesis. I may not graduate on time.

I ended up not graduating in the spring of 2012. I didn't finish my thesis on time and had to finish it over the summer. This prevented me from really looking for jobs and starting my life after college because I had this huge thing hanging over my head. I ended up doing the work that I should have done in the spring while I was supposed to be graduating having a good time. I eventually finished my thesis by the end of the summer, and now I can pursue my career 3 months after expected. Finally finishing and graduating is the greatest feeling ever. You want to do it on time and you want to do it worry free. I wish I could go back and just get all of the work done in the fall of 2011 because if I would have had everything planned out and ready for my thesis class, then it really wouldn't have been that hard or stressful. I made it much harder on myself than it had to be all because I thought I could get it done in a couple of weeks. THIS IS NOT THE CASE! You are going to run into problems and you are going to have to change things during the process of writing your thesis. It is time consuming and should be taken very seriously. I did not and paid the price.

Katelyn Klinck
Thesis Report

Personal Timeline:

I began work on my senior thesis knowing that I wanted to pick a topic relevant to today's current economic issues. I also had found my health and labor economics courses very interesting and knew that I would like to do something regarding those topics. I narrowed my possible topics down to three choices and researched literature on those topics. I found an overwhelming amount of information on the current economic crisis, the baby-boomer generation, and their impact on retirement over time. I was able to narrow in on this topic around the end of September.

I worked on my literature review, mainly gathering research, throughout the first semester. I was able to submit a written copy of my prospectus before winter break. During winter break I edited my prospectus based on Professor Naples comments and began to gather data. The data gathering took much longer than expected. I kept thinking of new variables that would benefit the model and then would at times find collinearity between the variables and would have to decide which to keep in the model. I started running correlations and graphing the data around the end of February in order to interpret the data I was gathering.

By the beginning of March I began to run my regressions, which led to more data gathering in order to correct for problems I saw when running the regressions. Based on the new data, and new findings, I had to re-run various regressions with different data. I also had to log the data and take the first difference in order to correct for heteroskedasticity and stationarity issues. I spent a total of 3 weeks running the regressions before I began to write the paper. It took 3 weeks because I had to re-learn some parts of econometrics and learn about stepwise regressions and STATA for the first time.

I began to write my paper around mid-March, finishing around the second week of April. Writing the paper pushed me to think more about the data and I realized problems I was having and what other data I would have to find and other tests I would have to run to fix these problems. Because of the problems I realized while writing the paper I only had some significant results to discuss at Ursinus, but I also had an idea of what was necessary to achieve my desired results. I was able to present these new findings at the celebration of student achievements at the end of April. After the celebration of student achievements presentation I had minimal writing and regressions to run. The few things I added to my paper were mainly based on suggestions from my presentation.

Knowing what I know now, I would attempt to do more of the writing for my prospectus (I focused a lot on researching and saved the writing for the last month) and gathering actual data (with correlations and graphs) in the first semester. My data gathering took the longest and left me less time to write the actual paper since it trickled into the second semester. I recommend that anyone working on their thesis should try to have a fully edited literature review and data with graphs and correlation matrixes by the end of the first semester so that you can focus on running regressions and writing the rest of the paper during the second semester.

Jordan Mojka
Thesis Overview

Unlike the majority of the other senior thesis students, I began my thesis at the very end of my junior year. While I would suggest beginning the thesis as early as possible, I needed to begin and complete my literature review (the portion of the thesis that is normally completed in Fall semester of the senior year) so I could write my thesis during the first semester of my senior year in order to be considered a full time student during my J&J co-op. As a result, my experience was very different than other students in that my experience was very independent and separated from the remainder of the seniors.

When I began the thesis my junior year, I originally wanted to focus on Economics of Law. However, after spending the next month and a half trying to find research on the topic that was experimental proved to be fruitless, I decided to change my topic to a project I worked on during a finance co-op I had in the beginning of that summer. During that project, I was asked to analyze the impact of different macroeconomic variables on the company's revenue. I found the topic extremely interesting and wanted to see if I could determine a revenue-forecasting model for a particular company. Thus, in the beginning of August I began to researching the topic. However, it was also extremely difficult to find research for my new topic. After a month of research, I found that no study had been done that created a revenue-forecasting model for a specific firm with macroeconomic variables. Instead, I needed to focus on studies that used macroeconomic variable to forecast macroeconomic events and focused on consumer consumption habits. I was able to finish my initial literature review by the very beginning of September.

Once my literature was finalized with Dr. Naples (approximately middle to end of September), I began looking for my data. I mainly use the Bloomberg terminal for my data in order to find Walmart's quarterly revenue and the multiple macroeconomic variables. However, finding my data was slightly difficult in Bloomberg simply because I had never used the terminal before. I was able to begin running a correlation matrix and regressions by the end of October. My regressions were relatively simple in that I only needed to correct for autocorrelation, which required me to use the "prais" command in Stata instead of regression command.

Since I wrote my thesis in the Fall semester, I had my entire thesis to present at both Ursinus and the Celebration of Student Achievement. While it was nice to be done with my thesis at that point and able to present the completed project, it would have also been nice to go back with the suggestions from both audiences and improved my thesis.

If I was able to re-do my thesis, I would have made myself more disciplined last summer when I was finding research for my literature review. It was extremely difficult managing two different internships/co-op's and working on the thesis after work, however, I believe that I could have completed the literature review earlier, and thus would have made more time to find data and run regressions, and thus determine a more complete model.

Overall, I would suggest that anyone working on his or her senior thesis should begin as quickly as possible simply because more time will enable them to look into additional findings that are presented by the final models.

AnnMarie Pino

Thesis Report

The topic I chose was on the non-genetic factors that influences the onset of type 1 diabetes. I did not take me long to come up with topic and I had this idea about one month into the fall semester. The literature review took approximately the rest of semester to research and write. Personally, I created a survey in order to collect data and this took about two months to construct and gain IRB approval. My survey was approved on February 16 after being submitted on February 13 in which I was given an expedited review. After spring break I had all of my data.

I began running regressions and writing about my theory and methods shortly after this. Regressions always take longer than you think so I would have liked more time to do this. Since my dependent variable was binary I ran a probit regression. Other tests to keep in mind: stepwise, heteroskedasticity, and correlation matrix. Although the time crunch was close, I had good data and results to present at the Ursinus conference. Afterwards I did run some additional tests before the second presentation at Student Achievement day. After this time I had a lot of editing and writing still left to do.

Some advice would be to think of your topic early and to have your literature review finished before Thanksgiving. Because of this, you can start collecting your data over winter break and then start running regressions at the beginning of the spring semester. Most importantly, utilize your professors. Talk to your past economic professors whose class you found interesting. All of them are more than happy to help and you never know what will trigger a topic idea. Be open to suggestions because once you are immersed in a large project such as this, it can be easy to miss things. Most importantly, choose something you are truly interested in because you are working on this for a year.

Jimmy Shaw

Thesis Report:

-Senior Thesis Prep: November and early December was when I spent most of my time research on my topic. I looked for articles and data. Since I was working with 3rd world countries, it was a nightmare trying to find economic data earlier than the late 1990s (Asian financial crisis). The World Bank had data 1.5 years behind the current date which drove me insane when talking about the most recent recession (2008).

-Senior Thesis (spring): My experience with writing and planning a thesis was terrible. There was little education on tips and tricks for writing a thesis. I was lucky in taking econometrics as a junior, but for these students taking it as a senior, they must of had a huge learning curve. I ran a typical regression and used vif and hettest commands.

-Presenting: I feel presenting at Ursinus and at Student Achievement should be optional. If the student wanted to share his/her work, they should have the option to. Also, if they are running behind in research and writing, this automatically causes the econ department frown upon those students. It should not be counted as a grade and honestly, we are seniors in college, we know how to field questions, we don't have time between finishing up other projects and finding jobs to be wasting our time talking to others in academia (no offense).

Overall: I understand the thesis is independent, but there needs to be some structure. I felt overwhelmed, did not feel valued as a student, and had anxiety because I was outside of my comfort zone with no one to be there for me.

I was given an incomplete and my final version was submitted the first week in June (grade still pending).

I hope this helps.

Alysson Wong
Thesis Report

Choosing a Topic

I began thinking about my thesis topic as soon as the fall semester began. I wasn't too sure what my topic was (obesity) until I began doing another project in my Global Public Health class which wasn't until mid-November. I then proceeded to do the basic Prospectus (Intro, Literature Review, and data sources) which I handed in, in early December. Looking through the literature I discovered the usual factors people associate with obesity (fast food consumption, not exercising, socioeconomic status, etc), and chose to see which the major contributor was by looking at several countries. Upon reading the prospectus, however, Dr. Naples advised me to expand my research and further my thesis since it may not be as challenging a feat as I anticipated. After exploring some more, I finally decided to include antibiotic consumption, the use of antibiotics on livestock, as well as confectionary consumption. I handed my complete Prospectus early January.

Data Collection and Running Regressions

Beginning the first week of the semester, I met with Dr. Naples once a week to discuss my data collection and the progress I was making. The data collection proved to be quite difficult, and much more time consuming than I had originally planned, especially for the more underdeveloped countries. The variables that proved to be the most difficult were pharmaceutical expenditures, motor vehicle consumption, and fast food prevalence. Eventually I was able to find the pharma data, excluding a few years, as well as one year's worth of car consumption which I chose to use as a fixed effect. For the fast food, however, I had decided to find the number of local and foreign fast food restaurants that existed in each country. Many of the local fast food restaurants did not provide their financial statements, so I was unable to measure the fast food consumption through the businesses' performance. Since there was not an exact list I had to go to each restaurant's website and count them individually using the restaurant locators. By the last week of March, I was only halfway through the number of U.S. restaurants and had to abandon the variable altogether.

Since the Ursinus Conference was fast approaching, I only had a little over two weeks to start running regressions, writing my thesis, and preparing for the conference. I ran into many complications with my results due to high correlations among the data, heteroskedasticity, autocorrelation, etc. Also, because I had panel data and due to the statistical software I was using, STATA, I had to manipulate the format of my input data that took up a lot of time; apparently the program does not like repeated years which I had due to my panel data. Due to these obstacles I had to transform the data several times (separate the data run regressions on individual effects, log the data, and take the difference in), and was not completely finished for the Ursinus Conference; though I was able to present my current results and future plans for the thesis.

After Ursinus Conference

The students and faculty that observed my presentation at the conference were extremely helpful, and provided valuable feedback on how to analyze my results. I did several more regressions following the conference and by Student Achievement day, I had completed the written part of my thesis.

What I would have done differently

While I was able to complete everything in a timely fashion, there are a lot of things I would have done differently that would have saved me a lot of time and energy. For one, I wish I could have come up with a finalized topic during the fall semester and began the data collection before Winter Break. Also, if I could have met with Dr. Naples a few times during the fall semester it would have sped up the whole process since she was an extremely valuable resource. If I had been able to do this, then I may have been able to focus solely on running regressions during the spring semester. This would have allowed me to present a lot more results, if not my complete thesis at the conference. Also, I should have done a bit more research on STATA commands before I started running regressions. That would have been much easier than searching for commands as I was working; I found looking them up quite time consuming. Overall, I just wish I had been more conscious of my time management since the spring semester comes to a close faster than one would think.