How to write an M Phil Thesis
Prof. Bronwyn H. Hall
Hilary Term 2001

Outline
1. What is a thesis?
2. How to find a topic
3. Research
4. The structure and content of a thesis
5. Primer on tables and figures

What is a thesis?
1. A proposition stated or put forward for consideration, especially one to be discussed and proved or to be maintained against objections.
2. A dissertation on a particular subject in which one has done original research, as one presented by a candidate for a diploma or degree.
Ideally, a thesis is a statement of an idea that you believe to be true accompanied by a clear presentation of the evidence to support it.

What is a thesis?
Often starts with a question:
– Why has wage inequality increased in recently?
– Is efficiency wage theory a good description of reality?
– Does foreign direct investment increase the technological capacity of developing countries?
The answer to the question becomes your thesis, in the first sense of the word.

What is a thesis? - examples
1. The European Union is (is not) an optimal currency union.
   – Review prior literature on currency unions
   – Define an optimal currency union
   – Collect some monetary data on European countries
   – Use the data to confirm or deny the assertion

What is a thesis? - examples
2. Research joint ventures can have positive or negative welfare effects.
   – Review prior theoretical literature.
   – Construct one or more models of firms engaged in research joint ventures.
   – Using your model, show that they can increase research toward the socially optimally level because some externalities are internalized.
   – Also show that they may facilitate collusion among the partners.
What is a thesis? - examples

3. Wage inequality in the UK has increased as a result of technological change.
   – Review prior evidence.
   – Lay out the reasons why wage inequality might increase (globalization, demographics, changes in educational policy, technical change, capital-skill complementarity, etc.)
   – Collect some data and investigate which of the reasons might be the cause (could be more than one).

How to find a topic

• What interests you? Is there a loose end in something you have learned that intrigues you?
• Many questions in economics are hard to answer and the work to date may have flaws (it helps to critique some of the papers you have read as you read them).
• Can you improve on an existing approach, using new methods or new data?
• Talk to economists in your field of interest to see if they have suggestions – it is helpful to have thought a bit yourself about the problem before going to see people. One of the economists in your field will probably be a good supervisor (see the departmental list of specialties for help).

How to find a topic

• More thoughts on getting ideas:
  – Read recent journal issues to see the kind of work being done in the area you are interested in.
  – Alternatively, you may find a question arising from reading the business/economics press – e.g., the Economist or the Financial Times.
  – Browse working paper abstracts on the web (see research resources listed later).

Research

• Literature search – use the web
  – www.nber.org for recent working papers from NBER.
  – Econlit at Oxford (OLIS) for older papers – easy to search abstracts for topics that interest you.
  – www.ssrn.com for recent working papers on all topics.
• Keep a list of references from the beginning of your project, preferably in a word-processed document – this will save time at the end.

Research

Make an outline – it is a good road map for the gaps that need filling, even if you have to change it as you go along.

It is also helpful to start writing before you finish the research:
  – It will clarify your thinking on the topic.
  – Something to give a supervisor to help them see what you are doing and where the problems are.

Research (empirical)

Your most difficult task:
• What kind of an “experiment” would answer the question you are asking?
• Do you have or can you get data that will enable you to have a “quasi-experiment” that is similar?
Examples:
• Returns to training in the armed forces (Angrist)
• Efficiency wage tests – associate productivity and wages across regions or industries?
Contents of an empirical thesis

1. Statement of thesis – why is this interesting and what is your contribution to the problem?
2. Previous research - literature survey (not a catalog).
3. Some modeling of the problem; possibly a discussion of the econometric method(s) to be used.
4. Description of the data used and how collected, including tables and/or figures of summary statistics.
5. Estimation results, analysis and discussion.
6. Some conclusions.

This is just a rough guide, and your thesis may take on a different structure, although most of these elements will be necessary in one form or another.

Contents of an empirical thesis

Theory: The role of economic theory in an empirical thesis is to provide a framework in which to think about the problem. Only rarely will you conduct an exact test of a very precisely defined model.

– Sometimes part of your contribution will be a new model – in this case spend some time describing it in the text.

– Other times an exact mathematical model will not be necessary or will be too restrictive. In this case the theory is just a guide as to what are the independent and what are the dependent variables.

Contents of an empirical thesis

Econometrics:
• Use appropriate methods – does your data satisfy the assumption necessary for the method to yield “correct” estimates. Why might it not?
• Conduct and report diagnostic and specification testing for your model.
• Explain how you performed the estimation clearly enough so the reader can understand. Reference the software used, including version number.
• Often the choice of econometric method is central to the empirical findings of your thesis and its ultimate quality.

Data:
• Many sources of data: OECD, IMF, various employment and labor force series in the UK. See also individual researcher websites (such as mine for patent data, firm data, especially R&D; the NBER for 4 digit manufacturing series, Summers-Heston growth data, etc.).
• It can be difficult to find exactly what you want – be creative and willing to cut corners if you have to.
• Some research starts with the discovery of an interesting new dataset (Babylonian prices?)

Primer on tables and figures

Warning: LaTeX, admirable though it is for typesetting mathematical formulas, is not good for producing tables. Use Excel or another alternative for this task.

Ideally a reader of your tables and figures should not have to refer to the text to understand them. Keep the following in mind:
• Not too many significant digits (usually 3 or 4)
• No E (exponential) notation for numbers – re-scale the data.
• Avoid strange acronyms for variable names – use short English descriptions if at all possible.
• Clearly report the sample size and some diagnostics to reassure us that regressions or other estimations are well-specified.
• Be clear about the choice of dependent variables, the units of the variables and the method of estimation.
• Report standard errors of the coefficient estimates (avoid reporting t-statistics or p-values, especially the latter).

Table 6a

Testing for Nonstationarity Using SUR

<table>
<thead>
<tr>
<th>Method</th>
<th>DF w/ Coeff</th>
<th>DF w/ Mean</th>
<th>DF w/ Mean and Coeff</th>
<th>DF w/ Mean and Intercept</th>
<th>DF w/ Mean and Intercept and Coeff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random walk</td>
<td>0.0003</td>
<td>0.00</td>
<td>0.17</td>
<td>0.15</td>
<td>0.06</td>
</tr>
<tr>
<td>Fixed effect</td>
<td>0.0003</td>
<td>0.03</td>
<td>0.09</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean and Coeff</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean and Intercept</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean and Intercept and Coeff</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
</tbody>
</table>

In the first row, this is the size of a one-tailed t-test for rho<1 with nominal size 0.05. In the other rows, it is the empirical probability of rejection by such a test. The data are simulated using parameters derived from the first and second moments of the process for log R&D in the United States. The model estimated is \( y(i,t) = \alpha(t) + \delta(\rho) y(i,t-1) + \epsilon(i,t) \). The method of estimation is seemingly unrelated regression, allowing for correlation across time.

Table 6a

Testing for Nonstationarity Using SUR

<table>
<thead>
<tr>
<th>Method</th>
<th>DF w/ Coeff</th>
<th>DF w/ Mean</th>
<th>DF w/ Mean and Coeff</th>
<th>DF w/ Mean and Intercept</th>
<th>DF w/ Mean and Intercept and Coeff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random walk</td>
<td>0.0003</td>
<td>0.00</td>
<td>0.17</td>
<td>0.15</td>
<td>0.06</td>
</tr>
<tr>
<td>Fixed effect</td>
<td>0.0003</td>
<td>0.03</td>
<td>0.09</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean and Coeff</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean and Intercept</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>AR(1) w/ Mean and Intercept and Coeff</td>
<td>0.0016</td>
<td>0.00</td>
<td>0.15</td>
<td>0.09</td>
<td>0.08</td>
</tr>
</tbody>
</table>

In the first row, this is the size of a one-tailed t-test for rho<1 with nominal size 0.05. In the other rows, it is the empirical probability of rejection by such a test. The data are simulated using parameters derived from the first and second moments of the process for log R&D in the United States. The model estimated is \( y(i,t) = \alpha(t) + \delta(\rho) y(i,t-1) + \epsilon(i,t) \). The method of estimation is seemingly unrelated regression, allowing for correlation across time.

Table 3
Effect of Adding Patents and Citations to R&D Regression
U.S. Manufacturing Firms (Cleaned Sample) - 1979-88, 19,628 firm-years
Independent Variable: R&D Stock w/ Dependent Variances - Replication of Table 1

<table>
<thead>
<tr>
<th>Equation</th>
<th>R&amp;D Stock/Assets</th>
<th>Non-R&amp;D Stock/Assets</th>
<th>Dependent Variable</th>
<th>R-squared</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>1.112 (0.043)</td>
<td>1.109 (0.043)</td>
<td>1.116 (0.043)</td>
<td>0.206</td>
<td>0.710</td>
</tr>
<tr>
<td>Base R&amp;D</td>
<td>1.139 (0.045)</td>
<td>1.137 (0.045)</td>
<td>1.146 (0.045)</td>
<td>0.211</td>
<td>0.708</td>
</tr>
<tr>
<td>Base Cit</td>
<td>1.189 (0.052)</td>
<td>1.187 (0.052)</td>
<td>1.199 (0.052)</td>
<td>0.228</td>
<td>0.701</td>
</tr>
<tr>
<td>Base R&amp;D Cit</td>
<td>1.159 (0.046)</td>
<td>1.157 (0.046)</td>
<td>1.168 (0.046)</td>
<td>0.216</td>
<td>0.706</td>
</tr>
</tbody>
</table>

Heteroskedastic-consistent standard errors in parentheses.
All equations include year dummies.
Citation stocks are patent stocks weighted by all the cites they received before 1994 plus an estimate of post-1993 cites, depreciated as of the patent date.

Table 3
U.S. Manufacturing Firms (Cleaned Sample) - 1979-88, 19,628 firm-years
Nonlinear Model with Dependent Variable = log of Tobin's q
Effect of Adding Patents and Citations to R&D Regression

Figure 1
US Manufacturing - Cleaned Sample - 4,846 Firms

- % with R&D Stock
- % with Patent Stocks
- % with Cite-wtd Patents
- % with Cite Stocks