

Doing an Applied Project

In doing your applied project you will be expected to:

- Select a question of interest and relevance to you
- Review very briefly the literature on that topic
- Discuss the methodology you are employing
- Collect data (usually secondary) relevant to the issue
- Prepare the data for analysis
- Analyse the data using some technique
- Reach some conclusions

These notes provide some advice on how to undertake a piece of applied work using the skills learnt in the Econometrics and Research Methods courses. This can be complemented by Appendix A in Intriligator, Bodkin and Hsiao (1996).

1. Starting

By the end of the first semester courses you should be able to run regressions and interpret the results. The applied exercises you undertook were designed to provide you with the required skills.

You should give some thought to the topic you would like to do before the beginning of the second semester. Once the course starts you will be asked to provide a brief proposal, describing the topic, the general form of the model and the data to be used. Surgeries to provide advice and guidance will be given in the second and third weeks and you will be assigned a member of staff to provide supervision. You will be required to make a short presentation of your work in week seven.

2. Topic

You can choose any topic. Useful sources are:

- an interesting dataset
- your job if you are a part time student
- your optional subject if you are full time
- published articles
- members of staff
- overseas students might choose topics using data from their own country.

Do not put a lot of effort into the theory until you have the data and avoid choosing a topic that is too ambitious. Also, do not worry about whether the results you get are good or bad, you will be assessed on how well you analyse a problem, not on your final result.

An important way of developing your project is to replicate a published paper. Applying the researcher's model to the same data, updating the analysis, and then maybe attempting to improve upon the model. Alternatively, you could use the data for a different country, industry or time period. It is often very difficult to get exactly the same results as those published. See Dewald et al (1986).

Read empirical economics papers and try to copy their style. Have a look through some recent journals.

- American Economic Review
- Economic Journal
- National Institute Review
- International Review of Applied Economics
- Applied Economics

are good places to look for empirical papers. Appendix A in Intriligator, Bodkin and Hsiao (1996) gives more. If you want to find a paper or book on a particular topic look in the Journal of Economic Literature, which classifies them by subject.

Do not agonise for too long over choosing a topic and once you have chosen a topic and collected the data do not be tempted to change.

The topics should involve explaining at least one variable by some others. Some examples might be standard economic relationships:

- explaining a country's imports by the level of demand, GDP; domestic process; import prices; tariffs; exchange rates.

- explaining consumption by income, inflation and wealth.

- explaining money demand by income and interest rates.

Or non-economic relationships:

- explaining attendance at football matches by ticket price, the quality of the teams, hooliganism, and the weather.

- explaining carbon dioxide concentrations by industrial production and sea temperature

- explaining crime rates by unemployment and economic factors.

3.Data

Finding the appropriate data can be the most difficult part of the project. You should check that the data is available before deciding on a topic and make sure you are clear what the data is. You can use time series data, cross section data (observations at one moment in time over countries, regions, families, etc...), or panel data which combines time series and cross section.

Make sure there are enough observations and variables. The sample size can be important in determining the techniques you can use and the precision of our results. Aim to have at least 30 observations for annual or cross section data; more for quarterly or monthly data. Unless you have experience of large data sets, or can get help in handling them, do not go much above 100 observations.

Make sure you know the exact definition of your data and what they are measuring. Terms like income and prices are not acceptable as they give little information. The sort of questions you should consider are:

Are the data current or constant price?

What is the base year?

What is the coverage (Net or Gross, Domestic UK or GB)?

Are they seasonally adjusted?

Have the definitions changed over the sample period?

If it is constructed data, how was this done?

Is the data based on a sample?

You may have to do a lot of work to make the data useable or comparable. If so provide information on this work, possibly in the form of a data appendix, so that you can be given credit.

You may have to adjust the data in various ways to deal with missing observations, to splice series on different bases, or to convert them into a different currency. Published data are not infallible, so always be on the look out for possible mistakes.

You should know something about the relevant history and institutions, such as important events, like strikes, wars, or changes of government.

Once you have loaded the data onto the computer, you should conduct a descriptive analysis. Print the data out and check carefully for typing errors. Plot the data and note the distinctive features such as trends, temporal dependencies, seasonality, unusual observations, etc. Calculate the means, variances of your variables and their correlations.

Repeat this process after you transform the data in any way:

- growth rates or ratios (the savings rate, the velocity of circulation, the share of profits) are often more informative because they are not dominated by trends.

- logarithmic transformations are often used in economic models as the coefficients can then be interpreted as elasticities; the change in the logarithm is approximately equal to the growth rate; variances are more likely to be constant; and many interesting economic hypotheses can be expressed as linear restrictions in logarithmic models.

The introduction to the article by Hendry on house prices in Hendry and Wallis (1984) is a good model for this descriptive analysis. As part of the data description for time series you should check the order of integration of the variables and whether they are co-integrated.

Where there are a number of possible measures for a series use all of them and try to decide which is best. You can report this in the project. e.g. Do wages respond more to the consumer price index or the retail price index? Do not decide a priori, test and find out.

Keep detailed notes on the sources of data and anything you do to them. Its easy to forget at a later stage. You might end up with some adjustments you made and forgot about dominating your results.

Keep at least two backed up copies of your data on separate discs, stored separately. There are many ways of loosing or corrupting disks and it can be a lot of work typing the data in again.

Sources

For UK data:

- Economic Trends Annual Supplement: is the best single source
- The Blue Book (National Income and Expenditure)
- The Annual Abstract of Statistics
- Financial Statistics
- Employment Gazette
- Family Expenditure Survey
- Census of Production
- Key Data a CSO publication provides a good introduction to what data is available and will give the relevant specialist publication.

For international data:

- World Development Report: published by the World Bank
- OECD Main Economic Indicators
- International Financial Statistics: Published by IMF

For the US:

Economic Report of the President: has an appendix giving the main economic aggregates.

Financial data:

Datastream: provides various financial and company accounts information and is available in the library. It also has some coverage of international macroeconomic and financial data. Similar but more limited data sources are available in the Hendon library.

There are also a number of publications which present data, such as:

- Maddison (1982) which gives long runs of historical data for the main capitalist countries.
- Berndt (1991) provides data on a floppy disc.

NB CSO, OECD and Eurostat data are available on disk if required. It is still a good idea to study the publications to make sure the data is what you think it is and to check on any idiosyncrasies

4. Analysis

Conduct your research with the final written project in mind. Make sure you can answer all the questions posed in the next section. Write fairly detailed notes of what you are doing and what your results are as you go along. It is very easy to forget what you did and be left with a vast pile of incomprehensible printout.

Try and organise your investigations around a few central questions. This will allow you to structure your specification search. The implicit methodology in these notes is broadly the Hendry/Spanos approach. This is not the only way to structure the analysis but it does provide a convenient framework.

5. Writing Up

Leave plenty of time for writing up, this tends to be the weakest link. In exercises like this students who can do all the rest, often fail to describe what they have done and what it all means. It should read like a good empirical paper in an economics journal, not a piece of autobiography. Read some more empirical papers and copy their style.

You are expected to type the project before submitting it, though equations can be handwritten. **Do not submit computer print outs.** You should process and digest the relevant information from them and report it in the project. Do not just transcribe the results of running dozens of regressions. Try to structure the interpretation of the results; pose questions and explain how the regressions provide answers to them. As you write up you are bound to think of something else you need to do. So start writing up early, don't leave it till the last minute.

Write about all the things the reader does not know and will need to know to understand what you have done. Do not copy large chunks of econometrics/statistics textbooks. The reader will know most of that, just give a reference.

Attempt the impossible: try to make it lively and interesting.

Introduction: Don't jump in at the middle. Introduce the subject, give some background information and refer to any relevant literature. Then explain the questions you are going to try to answer, or the problems you are going to solve and why they are interesting. Say how your project differs from other work.

Theory: Set out the economic theory and use it to specify a model. Wallis (1979) and Berndt (1991) are good on the process of moving from economic theory to econometric models. Discuss the economic interpretations of the parameters (elasticities, marginal propensities, long and short run effects, etc). Set out any a priori expectations about signs and magnitudes of the parameters. Set out any hypotheses to be tested: constant returns to scale; homogeneity; unit elasticities. Note any identities linking the data. Discuss any identification problems.

It is often useful to think of economic theory as specifying: a long run equilibrium relationship; an adjustment process; an expectations formation process. Discuss each of these separately. Think about the time series structure of the data. You will get into trouble if you try to explain a stationary variable just by a single, highly trended, variable. Check that orders of integration match.

Remember that some theories imply that variables should follow random walks, so the basic theoretical model is that the change in the variable is a 'white noise' error, unpredictable from earlier information.

Data: Discuss the sources of the data, the exact definitions of the variables, the sample used for estimation, the correspondence of the data to the relevant theoretical concepts, the possible measurement errors etc. Describe the main features of the series, with graphs if necessary, and point out any peculiarities or outliers. Ask whether the series are stationary in levels or first differences.

Credit is given for data collection but you need to give information about it. If you have done a lot of work developing a new or unusual data set make sure you describe what you have done.

Statistical Model: Use the theoretical and the probabilistic structure of the data to choose a statistical model; linear regression model; dynamic linear regression; multivariate regression; vector autoregression; simultaneous equations model; etc. Discuss your choice of statistical model in terms of the assumptions it involves. You need to convince the reader that you have made an appropriate choice. It pays to start by trying both a simple model using just levels of the main variables (which you may want to treat as a cointegrating regression) and a fairly general model with lots of lags and variables.

Estimation and Misspecification Testing: Estimate the statistical parameters of interest and test the validity of the assumptions underlying the statistical model (no serial correlation, linear functional form, homoscedasticity, normality, constant parameters, etc). If any of the assumptions are rejected you should respecify the model and try again. With luck you should get a "well defined statistical

model" that passes all the misspecification tests. Report the results briefly. How you lay out and present the results is very important. Try to copy articles in the literature.

Specification Testing and Interpretation: When you have a "well defined statistical model", then you can proceed to reparameterise/restrict the statistical model in order to construct an empirical econometric model. This involves testing economic hypotheses (eg homogeneity in prices, constant returns to scale), calculating the economic parameters of interest (eg long run solutions, elasticities), and interpreting the adjustment process (eg error correction, common factors).

Finally, evaluate your chosen empirical econometric model in the light of the original theoretical model, the estimated theoretical parameters of interest, and how your results compare with other published estimates. If you were unable to find a well defined statistical model go through this stage anyway, but point out that your results may be less reliable because of the possible misspecifications.

Conclusions: Explain the significance of the results and how they relate to the original questions or problems posed in the introduction. What is their relevance for practical questions of policy, forecasting, business? Are they consistent with theory and with institutional and historical information you might have? Is the model statistically adequate in representing the data.

References: Provide a list of works cited at the end, with references in the text of the form Berndt (1991).

References:

- Berndt ER (1991) "The Practice of Econometrics", Addison-Wesley.
- Dewald WG, Thursby JG and RG Anderson (1986) Replication in Empirical Economics, American Economic Review, September, p587-603.
- Hendry DF and Wallis KF (eds) (1984) Econometrics and Quantitative Economics, Basil Blackwell.
- Intriligator, Bodkin and Hsiao (1996) "Econometric Models, Techniques and Applications"
- Maddison A (1982) Phases of Capitalist Development, Oxford.
- Wallis KF (1979) Topics in Applied Econometrics, Basil Blackwell.