Scottish Graduate Programme in Economics
SESSION 2016-17

ECONOMETRICS PROJECT OUTLINE
(version of 15.10.2016)

Mark Schaffer (M.E.Schaffer@hw.ac.uk**)**

INTRODUCTION

You are required to write a group project demonstrating your ability to use econometric techniques in a practical context. This is a stand-alone 10 credit course.

The project is an empirical teamwork project. Its aim is to test your ability to (i) identify precise hypotheses that can be tested empirically, (ii) familiarise yourself with the relevant literature, (iii) work with a dataset which enables you to test your hypotheses, (iv) use Stata (or other) software to analyse the data, run the relevant regressions and test the hypotheses in question, (v) present your results to an audience, and (vi) write up the whole exercise as a group project.

Relevant dates:

- Noon, Monday 5 December: Final choice of topic and group membership.
- 4pm, Tuesday 10 January: Submission of presentation slides for SGPE Conference in Crieff.
- 12-13 January: Presentation of research plans at SGPE Conference in Crieff.
- Noon, Monday 27 March (first Monday after end of options courses): Submission of final version.

Details of the above are provided below. The second half of this document is a short guide to how to write an econometrics project.

CHOICE OF TOPIC AND GROUP MEMBERSHIP

The deadline for choosing a topic and finalizing group membership is **noon, Monday 5 December** by email to sgpe@ed.ac.uk. Groups are meant to be self-selected and should consist of 3 or 4 students.

A list of potential project topics is provided separately. You may choose one from the list or propose your own topic (see below). Some topics on the list are more advanced than others and are flagged as such. More than one group may choose to pursue a particular topic; there will usually be plenty of scope for groups to pursue distinct research strategies. Also please note that some topics draw on time series econometrics

methods that you will not see in the lectures until the 2nd semester. That said, reading ahead on time series methods is not a bad idea!

Please note that some of the proposed topics are more limited in scope than others and may therefore look as if they constitute easier project topics. As a result, when choosing a topic you should bear in mind the purpose of this exercise and the fact that there is a clear trade-off involved. Your choice of topic and how you propose to approach it reflect your ambition and will therefore be rewarded: when you choose a more demanding or a broader topic you are giving us the opportunity to assess better your overall skills. On the other hand, since choosing a more limited topic does not enable you to exercise all your skills, we would normally expect you to demonstrate a relatively higher quality of work in applying a narrower range of skills.

PROPOSING A TOPIC OF YOUR OWN

You may also decide to do a topic outside the list provided. If so, the topic will need to be approved by Mark Schaffer (m.e.schaffer@hw.ac.uk).

Should you decide to propose your own project topic, you will need to submit the following along with the list of group members:

1. Project title
2. A short description (approximately 1 page, can be less) of the project.
3. A list of references.
4. A brief statement about data availability.
5. A backup project from the "official" list of projects in case your proposed project topic is not approved.

It is absolutely essential that your project proposal is based on existing literature and research. The standard approach would be similar to that used in the "official" list of project topics: identify an empirical question, identify the papers that have explored this question using econometric techniques that you have studied or can learn, and replicate/extend this work and/or apply these methods to a different dataset.

Good places to get ideas for your own projects are journals/textbooks/researchers that make their datasets available along with their analysis. For example, the *Journal of Applied Econometrics* maintains a data archive at http://qed.econ.queensu.ca/jae/ for most papers they have published since 1994. Some researchers are very conscientious about putting their datasets online along with their papers: a good example is Andrew Rose at Berkeley (http://faculty.haas.berkeley.edu/arose/RecRes.htm). (In fact, if you look at the links to his "Surprising Similarities" paper on his website you will see a reference to 4 economists who spotted some issues with the econometrics in the paper. These were the members of a 2013-14 SGPE econometrics project group!) Many econometrics textbooks have empirical exercises based on published papers that offer an easy way in to the papers (Hayashi, Asteriou & Hall, Becketti, Greene and many others.)

PRESENTATIONS AT THE SGPE CONFERENCE AT CRIEFF

Groups will be required to do very short presentations of their research plans at the Crieff Hydro conference to be held on the **12th and 13th of January**. The material presented will typically consist of a literature review, a summary of your research strategy, and a statement of progress to date. It is *not* necessary for you to have concrete econometric results to present at Crieff. What *is* necessary by then is that you have a clear plan of what it is you are trying to do – including the econometric methodology to be used and the data required – and how you intend to do it. Each group will be assigned a 20 minute slot and a discussant. You should prepare a 10 minute presentation and be ready to answer questions for another 5-10 minutes afterwards.

The presentation and slides will have a 10% weight in your project mark (slides 5%, presentation 5%). The slides must be submitted by email to sgpe@ed.ac.uk (cc to M.E.Schaffer@hw.ac.uk) by **4pm on Tuesday, 5 January**. You must use the submitted version of your slides during the presentation.

PREPARATION AND SUBMISSION OF THE PROJECT

The final written version of the Econometrics Project should not exceed 8,000 words excluding references and appendix material. Please note that this is a **maximum** length; high-quality projects that are also concise are very welcome!

The project must be presented in the format of a typical academic journal article. Either MS-Word or PDF format is acceptable. You should use Times New Roman 12 point font, setup your page as with A4 paper, 1 inch (2.5cm) margins from each side (top, bottom, left, right), 1.5 line spacing, justified margins with footnote in the same page rather than endnotes. Use the same paragraph format throughout, title all your sections, tables and figures clearly and number all your pages, equations, tables and figures sequentially. Also, please note that you need to present your estimation results in the form of properly formatted summary tables; simply including unedited Stata or other statistical package output is **not** acceptable. These constraints are imposed so as to achieve a standard professional text format and you may be penalised (up to 10%) if you seriously violate them.

The submission deadline is noon on Monday, 27 March (the first Monday after the end of options teaching).

On or before this deadline, each group is required to deliver the following to the SGPE office:

- One hard copy of the project.
- A hard copy "declaration of own work" signed by each member of the group.
- A completed form (available on Learn) summarising how the team has worked as a group and any problems encountered.

- By email to the SGPE office (sgpe@ed.ac.uk) with the following three files as attachments:
    - The project in the form of a Word or PDF document (this should match exactly the hard copy of the project).
    - An Appendix in the form of a file containing appropriately annotated Stata (or other statistical package) output.
    - Your final Stata data file (with raw data and all necessary transformations included).
- Each group must also nominate one person who will submit the project to the TurnItIn (plagiarism) website for checking on the very same day. Details of the procedure for submitting a TurnItIn report will be circulated separately. The main aim is to show you how powerful this tool is for detecting copied paragraphs etc. rather than to enable us to impose plagiarism penalties, but it serves the latter purpose nevertheless!

A BRIEF GUIDE TO WRITING AN EMPIRICAL PROJECT*

The main purpose of the exercise is to enable us to assess your ability to use a coherent and relevant data set and relevant econometric methods to estimate some parameters of interest based on a clear theoretical relationship and test some hypothesis regarding the theoretical restrictions that these parameters are expected to satisfy, and to interpret the evidence in the light of the underlying economic theory.

As an economics graduate, you may need to write reports in professional capacity that involve analysing economic data. However, depending on the topic, purpose and intended audience, the form of these reports can vary widely, so that there is no one correct format for an empirical paper. The intention of this discussion is to give you some idea of what you should keep in mind while carrying out the project and presenting your results.

The first thing worth stressing is that there are no right or wrong empirical results. Empirical results are what they are and you should not be disappointed if they do not show what you had hoped they would. In an ideal world, a researcher first comes up with a new theory and then carries out empirical work that supports this new theory in a statistically significant way. But the real world very rarely resembles this. Many things contribute to this: (i) estimates of the crucial parameters are either statistically insignificant or have the wrong sign; (ii) the model fails one or more specification tests (e.g., exogeneity); and many other similar problems.

In empirical work one faces these problems all the time. But these should not discourage you! Instead, you should always keep an open mind. A finding that a theory does not seem to work is just as scientifically valid as a finding that a theory does work. Furthermore, empirical results are often unclear or confusing. For instance, one statistical test might indicate one thing while another indicates the opposite, e.g., two alternative tests statistics may lead to opposite conclusions. Likewise, an explanatory variable that is

* This is a modified version of an earlier document prepared by Hassan Molana.

significant in one specification of the regression equation might be insignificant in another specification. There is nothing you can do about this, except to report your results honestly and try (if possible) to understand why such conflicts or confusions are occurring.

The important thing is that your results are obtained from a sound econometric approach and appropriate data and are robust. You should not be tempted to do anything that is even slightly dishonest in order to show that your results match what you expected a priori. For instance, sometimes in empirical research it is tempting to run a large number of regressions with many different explanatory variables in the hope that one regression gives the result one wants. While this can be informative in terms of exploring the data in detail and from a number of different angles, it is dishonest if the researcher presents only the regression that supports a particular theory and not the other regressions that discredit it, as this will be an act of misleading the reader intentionally. Always avoid the temptation to misrepresent your results!

On the issue of how results should be presented, one cannot stress enough the importance of **clarity and brevity**. The end user who reads the paper is usually busy and does not want to spend a lot of time reading long, poorly organised and verbose reports. One key skill that you are expected to develop is the way you select the most pertinent result. You may have many different estimates and statistical tests from your various regression runs but you cannot just present everything – you need to be selective. An important skill of project writing is the ability to sift through the information you have generated and select the subset which is most important for the occasion.

In conclusion, select only the most important information for presentation, justify your selection, and report the results honestly and openly so that they can be replicated if necessary by other researchers. With this in mind, we provide below the most common elements of economic reports as a guideline for future empirical work. This is a general guideline on writing an empirical paper or project; in your project you do not have to include every section that is mentioned here. These are included here to give as thorough a picture as possible what an empirical paper involves. The sections that a typical project should include and their brief description are listed in the relevant order.

(1) Cover page

This provides the title, names of the authors and their affiliations, a short abstract (see next point), a list of the most relevant keywords, the Journal of Economic Literature Classification (see http://www.aeaweb.org/journal/jel_class_system.php)

(2) Abstract

You must submit a structured abstract of up to 300 words. For guidance, see *Nature*'s "How to construct a *Nature* summary paragraph" document at www.nature.com/nature/authors/gta/Letter_bold_para.doc (also reproduced at the end of this document).

(3) Introduction

Most reports begin with an introduction that briefly motivates and describes the issue being studied and summarises the main findings. The introduction should be written in simple non-technical language, with statistical and economic jargon kept to a minimum. A reader who is not an expert in the field should be able to read and understand the general issues and findings of the report or paper. A single page should suffice.

(4) Literature review

This should summarize related work that others have done. It should list and briefly describe how other studies have examined the related issues and how their findings relate to what is undertaken in the project. This gives you an opportunity to strengthen the motivation for your study. Two pages at most are probably needed.

(5) Economic theory

This section provides a description of the theory underling the phenomenon being studied. In academic papers, this section can be quite technical and usually involves discussing and comparing alternative theoretical models that can be employed in the study. Academic papers may provide detailed mathematical descriptions but policy reports usually do not need to include such details. This section allows you to describe the economic intuition underlying the question you are focussing on, and to outline any issues that affect the theoretical specification of your model in more detail. The discussion in this section leads to the identification of the model or relationships which are to be estimated. Two pages at most are probably needed.

(6) Data

In this section you should describe your dataset, give a clear explanation of the sources from which you taken them, and provide a good discussion of the measurement issues and why the series chosen are good proxies for the respective variables. You need to show that you know your data and their characteristics and summarise these clearly for the reader. Apart from the source and nature of the data (i.e., cross section, panel, survey, time series, balanced or unbalanced, frequency, existence of any breaks, etc.), you also need ideally to provide a summary of the stylised facts. This would be in the form of a discussion of the features of the data illustrated by scatterplots or other graphs, followed by tables of descriptive statistics (e.g., mean, standard deviation, minimum/maximum of each variable, a correlation matrix, some sub-sample statistics, etc.). Sometimes it is more appropriate to present the latter details together with the main evidence in the relevant section (see below). Three pages ought to be enough, but be sure to prepare your graphs by editing them properly and reducing their size.

(7) Econometric methods and issues

In this section you should discuss how you wish to go about using your dataset and econometric methods to analyse the model empirically so as to find robust answers for the particular economic questions you are asking in your study/project. The exact form of this section will vary considerably, depending on the topic. For instance, you may want to discuss some or all of the following issues:

- which variables can be treated as exogenous;
- why a particular specification of the regression equation is preferable;
- why a certain variable is chosen as the dependent variable;
- what estimation method is relevant;
- what is the best choice of set of instruments;
- whether or not the series should be transformed in any way and if so why;
- etc.

In short, it is in this section that you should justify the choice of your regression equations and the econometric techniques. This is an important section of your project as it shows your understanding of the course. But you need to be very precise and concise in your argument, so try not to exceed 3 pages.

(8) Results

This is possibly the most important section in any report as it tells the reader what the empirical results are and what they imply. Here you should describe your empirical findings and discuss how they relate to the issues/questions/phenomena under investigation. You should provide two types of information:

- Economic information which stem from the underlying economic theory. These concern issues such as (i) a priori expectations about the sign and size of coefficients (e.g. elasticities, etc.); (ii) the nature of causality; etc.

- Statistical information in connection with the application of econometrics to the data. These include issues such as: (i) How robust are the regression results? (ii) Are all the assumptions made in choosing an estimation method met by the data, and if not why and what remedies are sought? (iii) Are the coefficient estimates significant? (iv) Do the results support the underlying theoretical hypotheses? (v) Do you keep or drop insignificant explanatory variables and why? (vi) How does the model fit the data? (vii) How does the model compare with the rival ones? Etc.

As far as the presentational aspects of this section is concerned, it is not uncommon, and indeed sometimes preferable, to begin with the discussion of the features of the data (rather than providing these in the data section, as explained above). This is followed by tables which provide the estimation results: coefficient estimates, standard errors, t-ratios (and/or p-values), and other summary statistics (sample size, R^2s, F-statistics, diagnostic checks on the residuals, etc.). When describing your results you should interpret them and relate them to economic theory. If your evidence does not support your theory – e.g., if an

economic hypothesis you postulated is rejected by your test(s), or if the coefficient estimates have the wrong sign – you should try to give a possible explanation as to why this might be so. For example, if a coefficient did not turn out significant, was it because (say) the data was too noisy, there was an inadequate number of data points, or, a rival hypothesis is better placed to explain the situation? You should end this section by drawing some conclusion regarding the implication of your findings. As this is the most important part of the project, you can devote, say, up to 8 pages to it.

(9) Conclusions

This section summarises the issues addressed in the paper and emphasises the most important empirical findings within the context of the findings you have discussed in the literature review section; the reader should be able to see how your results compare to others'. You should be brief, precise and coherent. Two pages should be adequate.

(10) Bibliography

You should include full bibliographic details of all the articles, books, data sources, etc. cited in the report.

(11) Appendices

This is the place to present the material – proofs, extra evidence, miscellaneous results, etc. – which are essential but whose presentation in the main text would generate distraction. You can also have specific appendices for Data, Figures, Tables, Algorithms, and similar material. These are referred to in the main text and readers can choose to go through them as they wish.

(12) Stata (or other statistical package) output and data

As noted above, you should also submit a file containing appropriately annotated Stata (or other statistical package) output along with your final Stata data file (with raw data and all necessary transformations included). Do not print these; submit them as files as instructed above.

nature

How to construct a *Nature* summary paragraph

Annotated example taken from *Nature* **435**, 114-118 (5 May 2005).

*One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.*

*Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.*

*One sentence clearly stating the **general problem** being addressed by this particular study.*

*One sentence summarising the main result (with the words "**here we show**" or their equivalent).*

*Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.*

*One or two sentences to put the results into a more **general context**.*

*Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline, may be included in the first paragraph if the editor considers that the accessibility of the paper is significantly enhanced by their inclusion. Under these circumstances, the length of the paragraph can be up to 300 words. (The above example is 190 words without the final section, and 250 words with it).*

During cell division, mitotic spindles are assembled by microtubule-based motor proteins[1, 2]. The bipolar organization of spindles is essential for proper segregation of chromosomes, and requires plus-end-directed homotetrameric motor proteins of the widely conserved kinesin-5 (BimC) family[3]. Hypotheses for bipolar spindle formation include the 'push−pull mitotic muscle' model, in which kinesin-5 and opposing motor proteins act between overlapping microtubules[2, 4, 5]. However, the precise roles of kinesin-5 during this process are unknown. Here we show that the vertebrate kinesin-5 Eg5 drives the sliding of microtubules depending on their relative orientation. We found in controlled *in vitro* assays that Eg5 has the remarkable capability of simultaneously moving at ~20 nm s^{-1} towards the plus-ends of each of the two microtubules it crosslinks. For anti-parallel microtubules, this results in relative sliding at ~40 nm s^{-1}, comparable to spindle pole separation rates *in vivo*[6]. Furthermore, we found that Eg5 can tether microtubule plus-ends, suggesting an additional microtubule-binding mode for Eg5. Our results demonstrate how members of the kinesin-5 family are likely to function in mitosis, pushing apart interpolar microtubules as well as recruiting microtubules into bundles that are subsequently polarized by relative sliding. We anticipate our assay to be a starting point for more sophisticated *in vitro* models of mitotic spindles. For example, the individual and combined action of multiple mitotic motors could be tested, including minus-end-directed motors opposing Eg5 motility. Furthermore, Eg5 inhibition is a major target of anti-cancer drug development, and a well-defined and quantitative assay for motor function will be relevant for such developments.