**Group Term-Paper Requirements Outline**

Spring 2021

Fred Engst

To get the most out of this class, you will need to do more than just read the text, do the homework, and study for the exams. You need to get your hands “dirty”, so to speak, by building an econometric model, so as to gain a deeper understanding of the issues involved in the “art” of econometrics. You are thus expected to write an econometric research paper with another student on a topic of your own choosing (as long as it is about economics, and employing methods learned in this class). This will involve:

1) identifying a problem that can be analyzed with econometric techniques,

2) formulation a model with testable hypothesis,

3) collecting data,

4) estimating your model using the techniques covered in this class, and

5) interpreting the results.

While published papers might not go to as much details as I am asking you to do here, these extra requirements are to demonstrate (to me) how much you have learned in this class. You should thus understand the reasons behind this outline, and follow it as much as you can. Points might be taken otherwise. (See the summary at the end of this outline.)

Your main paper should not be more than 6 pages long, and no more than 10 pages including all graphs, tables and data appendices. You might have an extra cover sheet with your names, title of your paper, etc., and number all pages except the cover page.

1. **Introduction / Review of the Literature**: (1/2 to 1 page)
   * In this section you **introduce** your topic, state why others might be interested in the topic, or your motivation for studying it.
   * You should then **summarize**, in a few paragraphs, what other researchers using **econometrics methods** have found on this topic (using econometric techniques), or a topic that is very closely related to yours, following author (year) or author (year, p.xxx) style. Be sure to provide references for them at the end of your paper. You must have at least one paper from a peer-reviewed journal.
   * Finally, **state** your own hypothesis.

**I can’t stress enough how important this review of literature is for your paper.** To identify a problem that can be analyzed with econometric technique, you really need to carefully study what other scholars have done. It will give you ideas about what are the issues involved in the area that you are interested. In other words, don’t reinvent the wheel. You are not likely to be the first person whom ever thought about this problem. Find out what others have done first, before getting too deeply into your own study.

You might use <http://scholar.google.com/> to search for papers on a topic of your interest, or look into the following list of possible sources that might give you ideas and/or data.

Research Paper in Economics <http://www.repec.org>,

Econlit <http://www.econlit.org>, Jstor <http://www.jstor.org>,

ProQuest <http://www.umi.com/proquest>, SSRN <http://www.ssrn.com/ern>,

NBER <http://www.nber.org>, IMF <http://www.imf.org>,

WTO <http://www.wto.org/>, World Bank <http://www.worldbank.org/>

1. **Model and Data:** (1 to 2 pages)
   * This is where you write down your model using the format shown for all equations as in your textbook[[1]](#footnote-2). Use the subscripts and Greek letters as indicated. Markdown the expected sign of each variable under its name. Use informative names for your variables instead of *X*1 to *X*k. Don’t be lazy, and make your reader a detective! For example, for cross sectional models (only if you can’t make it into a panel data model) use

log(*wagei*)=*β*0 + *β*1*educi* + *β*2*experi* + *β*3*femali* +*ui*

(+) (+) (-)

where *wage* is the hourly wage in dollars,

*educ* is the year of schooling a worker has,

*exper* is the years of working experience a worker has, and

*female* is an indicator variable if the person is a female, 0 otherwise.

and for time series models (only if you can’t make it into a panel data model), use

Δlog(*GDPt*) =*β*0 + *β*1Δ*intt* + *β*2Δlog(*M*2*t*)+ *β*3Δlog(*FDIt*)+ *ut*.

(?) (+) (+)

where *GDPt* is the Gross Domestic Production (million dollars) at time *t*,

*intt* is the interest rate (3-month Treasurer rate) at time *t*,

*M*2*t* is the money supply (million dollars) for time *t*, and

*FDIt* is the Foreign Direct Investment (million dollars) at time *t*.

Or, for panel data[[2]](#footnote-3) mode use

log(*wageit*)=*β*0 + *β*1*educit* + *β*2*experit* + *β*3*kidsit* +*ai* +*τt* +*uit*

(+) (+) (-)

where *wageit* is the hourly wage in dollars for *i*th worker at time *t*,

*educit* is the year of schooling the *i*th worker has at time *t*,

*experit* is the year of work experience the *i*th workerhas at time *t*,

*kidsit* is the number of kids the *i*th workerhas at time *t*.

* + You should put **variable definition**, their unit of measurements in a table for easy reference.
  + You then **explain** the underlying economic theory behind your variable selection and your hypothesized signs of the coefficients if you did not indicate under your equation.
  + You should pay close attention to the **unit of measurement** of your variables and **normalize** them so as to be compatible with each other and to reduce heteroskedasticity. For time series data, be sure to deflate them properly. For cross sectional data, be sure to think about the correct functional forms for the variables, and see if you want to match gross of one variable with gross, average, or percentage, etc. of another variable, so that the variables will be generally compatible (and will not be affected by entity merging of splitting), to reduce heteroskedasticity, as well as for the ease of interpretation.
  + You should have at least **one nonlinear functional form**,such as logs, polynomials or interactions, and justify your linear or nonlinear functional forms by a few scatter plots (included in the appendix).
  + Next, you **discuss** the **data sources** for your model (you **must use the latest available data**, and no dataset from any textbook is allowed for this paper). It should have **50 or more data points (over 100 for time series)** and **3 or more independent variables** (lags or powers of the same variable doesn’t count), and from more then one data source (or data table). Furthermore, if your model has a time series component you should keep as much data as the data series is available. Otherwise, you must justify why you are using only a selected section of the time series data.
  + If you collected the data yourself, be sure to document the data collection methodology carefully.
  + You should explain all data transformation, smoothing or any adjustment performed in this section.

1. **Analysis and Results**: (1 to 2 pages)
   * You should then present your regression results in standard format, as shown below. You should include estimated coefficients with associated (robust) standard errors[[3]](#footnote-4) (use ‘\*’, ‘\*\*’, or ‘\*\*\*’ to indicate significance of 10%, 5% or 1% level respectively), *R*2, , *F*-statistics, *D.W* -statistics (for time series only), as well as the sample size *n*. They should be independent of the software used. For example, using HTV dataset resulted in:

*=*0.174+0.116*educ+*0.034*exper+* 0.0083*motheduc+* 0.021*fatheduc*

(.196) (.0105)\*\*\* (.00688)\*\*\* (.0086) (.0063)\*\*\*

*n*=1230, *R*2=0.19, 2=0.188, *SSR*=.534, *F*=72.2

Do not report the results in *gretl,* EViews, STATA, SAS, *R*, or SPSS[[4]](#footnote-5) output format here (instead, you **must** include the original output without editing in the appendix, including the STATA or *R* commends, or SAS log).

* + Generally, you should only present your final model or models, and briefly describe how you arrived at this, leaving the detail of modeling steps to the appendix if need be[[5]](#footnote-6). If you do need to compare results from different regressions, it is best to present them in a tabular form, and be sure not to let your table run over multiple pages. If you must, it should have new headings to indicate such.
  + It is important that you present the regression results in laymen’s terms. You will need to **discuss** the following points in your narrative:
    1. State the significance of each of the independent variables in your model using *t*-test, or p-value (only if you did not mark the standard error with appropriate asterisks ‘\*’ as in the example above), and relating your findings to the hypothesis stated in section 2 above.
    2. Report any statistical test conducted for the variable and/or model selection, such as *F*-test for joint significance of a set of variables, serial correlation test, and the test for heteroskedasticity, etc. **This is one of the most important sections of your report**, for it is where you demonstrate your mastery of the statistical concepts learned in the class; demonstrate your ability to apply those concepts in actual situations. However, don’t conduct all tests that you can find, just to show that you can do them. Conducting irrelevant tests exposes your lack of understanding! And it might event herm your grade!
    3. Interpret the meaning of the transformed variables, if you have any.
    4. If you use intercept dummies or slop dummies, justify their usage.
    5. **Interpret** the meaning of the coefficient estimates in laymen’s terms[[6]](#footnote-7) and be specific about what is being held constant (for I want to see how much do you understand the results from your own model).
    6. Could there be bias in estimated coefficients due to omitted variables? What variables might have been omitted and what are the likely effects on the estimated coefficients?
    7. Are the model assumptions justified, such as E(*u*|*X*)=0, or E(*uit*|*Xit*)=0?
    8. If your model is a cross-sectional one, you should test for the presence of heteroskedasticity. If it is found to be significant, you should discuss its consequence and state any corrective measures that you’ve taken.
    9. If your model is a time series one, you should test and correct for the problem of nonstationarity, serial correlation, etc.

1. **Conclusions**: (1/2 page) You should summarize your findings here, such as

* What conclusions can you make based on your study.
* How is it compared to those you’ve found in your review of literature?
* Does your analysis support your initial hypotheses?
* Based on your model, what policy recommendation could you make, if any[[7]](#footnote-8)?
* What suggestion you might have for future research?

1. **Bibliography**: List all sources that you used in your project. It includes references you made in the review section of your paper, and any other written sources that you have consulted. List items alphabetically by the last name of the author or the name of the organization that produced the item (not its web address). List each item’s name (year), “Title of Article” /*Title of a book*, *Title of journal or magazine* volume, page number /city of publication: publisher. If you use Internet as your source, it should not be the only source.
2. **Appendices:** (1-2 pages) You should include in the appendices the following (some of the items asked here is to help your instructor to understand your model, not to be included for publication):
   1. Clear indication of the data source, so that others can replicate your results without difficulty.
   2. A table of the first 5 and last 5 observation from your dataset used for regression with clear labels for each column, even for panel data.
   3. A table of summary descriptive statistics with clear labeling for each variable and its unit of measurement. (It might help you detect some data errors)
   4. A few informative scatter plots of some of the key non-dummy independent variables verses the dependent variable used in your model. For panel data, you either plot the combined data, or plot only for a particular year.
   5. The key regression outputs from *gretl* or any other software you used.
   6. Any other model run or tests you’ve conducted.

**Note:** You will not be graded on how will your model performs. Instead, you will be evaluated by how well you use and explain the statistical test employed and how well you convey and articulate your ideas. Outstanding work in review of literature, data collection, using advanced models learned in this class, or thoughtful statistical tests might lead to **extra points**. Using color in your printout adds no points to your grade in most cases, and any distracting graphics might even harm your grade.

Remember your reader (that is me) don’t have a lot of time to read and grade your paper, so don’t make your reader a detective in finding out the nature of your data, the variables you’ve used (give meaningful and easy to understand names, for example, use GDP/Pop rather than GDPPC, or log(GDP) rather than LGDP), and what is the aim of your paper, etc.

**Grading Criterion (Tentative):**

**Introduction / Review of the Literature (10%)**

Introduction to your project / Motivation (how clear is the goal of your project) Literature review (how is your work compare to others)

**Model and Data (45%)**

Theoretical framework

Model formation

List of relevant variables (whether or not you can find them)

Data collection (effort)

Variable definition

Variable format (proper functional forms)

Expected sign

**Analysis and Results (35%)**

Presentation and interpretation of your regression result(s)

Stat tests (Heteroskedasticity, serial correlation, F-test, etc.)

Analysis of omitted variable bias

**Conclusions (5%)**

**Style and Presentation (5%)**

**Summary of major point deductions if this instruction was not followed:**

Plagiarism is **100** points off!

Missing review of literature is **20** points off!

Using **outdated data** is up to **20** points off.

**Independent variables** less then 3 is **10** points off for each missing variable.

**Observations** less then 50 is 1 point off for each <50, with a maximum of **5** points off.

Report the regression results in **original software** output format (including data sample) within the main report is **5** points off.

Missing **unedited** **original output** in the appendix is **5** points off.

Use **stepwise regression** is **5** points off.

**Dragging** your reader through all of your intermediate regression results is **5** points off.

Missing a section on **interpret** the meaning of the coefficient estimates is **5** points off.

**Page numbers** >10 is 1 point off for each page, with a maximum of **5** points off.

Making your reader a detective is **5** points off.

1. Unless you have a strong reason for it, you should avoid identity regression, such as

   *GDPt=β*0*+ β*1*Ct+ β*2*It+ β*3*Gt+ β*4*NXt+ut*, for which all of the coefficients should be 1 except for *β*0. It is not clear what purpose a model of this kind serves. But if you insist, you should test this hypothesis. [↑](#footnote-ref-2)
2. I strongly encourage you to make your time series or cross sectional model into a panel data model. Running a time series only model most likely requires more advance tools, while a simple cross sectional model most likely will suffer from omitted variable problem. [↑](#footnote-ref-3)
3. You **must** use heteroskedasticity and/or autocorrelation-robust standard error estimate for all your models. The reason for reporting the standard error and identify its significance level rather than the *t*-value directly is that you can also compute the confidence interval easily, where as reporting the *t-*value makes this task cumbersome. [↑](#footnote-ref-4)
4. Don’t use the stepwise regression! It is a mindless process that most likely resulted in a meaningless model. [↑](#footnote-ref-5)
5. Don’t drag the reader through all of your intermediate regression results in your main paper. Present only the meaningful and what you think is the correct model or models. [↑](#footnote-ref-6)
6. Use actual unit, such as yuan, 10,000 yuan, or percentage point, etc., not just the word “unit” itself. [↑](#footnote-ref-7)
7. You should avoid any assertions or opinions in your paper that are not supported by the data or your model; avoid any fashionable but meaningless sloganeering phrases that government officials tend to make in their public speeches, or in government publications. [↑](#footnote-ref-8)