计量经济学－课程论文撰写规范说明

● 选题：基于现实经济/金融/社会问题，自主选题、建立计量模型（建模过程，必须规范！），正文字数0.5~1.5万字；或者，找一篇公开发表的正式学术期刊论文，"更新数据"、模仿重建重估模型，但不能完全重抄别人的实证结果与解释。

● 采样：采集的数据，必须说明样本数据范围、数据来源、描述性统计数值特征等信息；另，若是时间序列数据，还必须将数据更新到可能的最新时点。数据量不大（30个样本点以内），请在文后附上数据；数据量较大（超过30个样本点），请在文后附上随机截取的10条数据。

● 特别说明：正文中"不允许"直接贴软件计算结果的贴图，必须按"公式法、或表格法"概要输出需要的信息与常见统计量、解释说明实证分析结果。

● 所有同学，各自独立完成、不以小组形式提交论文，无论是否被抽检、每人都要提交纸质论文（建议准备PPT、但不强制，若自告奋勇展示论文的同学、必须准备PPT；PPT均不用打印、也不用提交，仅课堂展示用）。

Carrying Out an Empirical Project

● **Goal:** Learn how to "complete a term project + write an empirical paper"

（部分内容是重复迭代的）

1. Posing a question

- ✓ *Knowing precisely* what question you want to answer is essential

- ✓ You can only collect your data if you exactly know your question

- ✓ You *can only* know whether you can complete your project in the allotted time if you know *whether the necessary data is available*

- ✓ You can only know if your research question is of interest to someone if you can precisely state it and discuss it with your classmates or instructor

Carrying Out an Empirical Project

- **Finding interesting research questions**

  - ✓ Choose *the area of economics/social sciences* you are interested in

  - ✓ Examples for typical research questions:
    - Labor Economics: Explaining wage differentials
    - Public Economics: Effect of taxes on economic activity
    - Education Economics: Effect of spending on school performance
    - Macroeconomics: Effect of investment on GNP growth

  - ✓ Look for published papers on the chosen topic using tools such as *EconLit, Google Scholar, the Journal of Economic Literature (JEL), Elsevier* etc.

Carrying Out an Empirical Project

● **Your research project should _add something new_**

  ✓ Add *a new variable* whose influence has not been studied before

  ✓ Expand economic questions to include *factors from other sciences*

  ✓ Study an existing question for *more recent data* (may be boring)

  ✓ Use *a new data set* or study a question for a different country

  ✓ Try out *new/alternative methods* to study an old question

  ✓ Find a completely new question (hard but possible)

  ✓ It helps if your research question is *policy relevant* or of local interest

Carrying Out an Empirical Project

2. Literature review

- ✓ A literature review is important to *place your paper into context*

- ✓ Use *online search services* to systematically search for literature, e.g.,

  Elsevier ScienceDirect, EBSCO, Emerald......

- ✓ When searching, think of *related topics* that may also be relevant

- ✓ A literature review can be part of the introduction or a separate section

Carrying Out an Empirical Project

3. Data collection

Most questions *can be addressed using alternative types of data* (pure cross-sections, repeated cross-sections, time series, panels)

● **Deciding on the appropriate data set**

✓ Many questions can in principle be studied using a single cross-section, but for a reasonable ceteris paribus analysis *one needs enough controls*

✓ Panel data provides more possibilities for convincing ceteris paribus analyses as one can control for time-invariant unobserved effects

✓ Examples for panel data sets: PSID (individuals), Compustat (firms)

✓ Panel data for cities, counties, states etc. are often *publicly available*

✓ Data sets are often available online, in *commercial database, official website, or journal archives*, or from *authors*

<u>间接数据来源扩充：</u>　【Chinese data】

- <u>国家统计局</u>：www.stats.gov.cn
- 中国人民银行：www.pbc.gov.cn
- 财政部：www.mof.gov.cn
- 国研网（国务院发展研究中心）：www.drcnet.com
- 中经网（国家信息中心）：www.cei.gov.cn
- 中宏网（国家发改委）：www.macrochina.com.cn
- <u>万得（Wind）金融终端</u>：www.wind.com.cn
- <u>国泰安Csmar数据库</u>：www.gtarsc.com
- CCER经济金融研究数据库：www.ccerdata.com
- EPS数据库：www.epsnet.com.cn
- CEIC数据库：www.ceicdata.com
- China Data Center (Univ. of Michigan)：chinadatacenter.org/newcdc/
- 中国调查数据库（CCER and Univ. of Michigan）：www.chinasurveycenter.org
- China Health and Nutrition Survey：www.cpc.unc.edu/projects/china
- China Human Capital Project (Univ. of Pennsylvanian)：www.ssc.upenn.edu/china/index.htm
- 经济发展论坛：www.fed.org.cn/
- 以及各类统计年鉴。

<u>间接数据来源扩充：</u>　【International data】

- Penn World Table：pwt.econ.upenn.edu/php_site/pwt_index.php
- <u>World Bank</u>：www.worldbank.org/
- IMF：www.imf.org
- United Nations：www.un.org
- OECD：www.oecd.org
- <u>US Federal Reserve Bank</u>：www.federalreserve.gov
- US Bureau of Labor Statistics：www.bls.gov
- US Bureau of Economic Analysis：www.bea.gov
- NBER online data：www.nber.com/data_index.html
- Economic Times Series Page：www.economagic.com
- <u>Thomson Data stream</u>：www.datastream.com
- Bloomberg Data：about.bloomberg.com/product_data.html
- <u>BVD财经系列数据库（欧洲）</u>：www.bvdinfo.com（包括 Bank scope）
- Inter-University Consortium for Political and Social Research：www.icpsr.umich.edu/icpsrweb/ICPSR
- Journal of Applied Econometrics Data Archive：www.econ.queensu.ca/jae
- KOF Index of Globalization：globalization.kof.ethz.ch/
- Charles Jones' datasets：www.stanford.edu/~chadj/datasets.html

8

Carrying Out an Empirical Project

● **Entering and storing your data**

- ✓ Data formats: 1) printed, 2) ASCII, 3) spreadsheet, 4) software specific

- ✓ *Important identifiers*: 1) observational unit, 2) time period

- ✓ Time series must be ordered according to time period

- ✓ Panel data are conveniently ordered as blocks of individual data

- ✓ It is always important to correctly identify and handle *missing values*

- ✓ *Nonnummerical data* also have to be handled with great care

- ✓ Software specific formats often provide good ways of documentation

Carrying Out an Empirical Project

● **Inspecting, cleaning, and summarizing your data**

  ✓ It is extremely important to *become familiar with* your data set

  ✓ Even data sets that were used before may contain *problems/errors*

  ✓ Look at individual entries/try to understand the structure of your data

  ✓ Understand *how missing values are coded* ; if they are coded as "999" or "-1", this can be extremely dangerous for your analysis; it is better to use *nonnummerical values* for missing values

  ✓ Understand the *units of measurement* of your variables

  ✓ Know whether your data is *real/nominal, seasonally adjusted/unadjusted*

  ✓ Check if means, std.dev., mins, and maxs of your data are *plausible*, and clean your data of implausible values and obvious coding errors

  ✓ When *making data transformations* (differencing, growth rates) make sure your data is correctly ordered and no wrong operations result; e.g., in a panel data set, be aware that the first observation of each cross-sectional unit has *no predecessor*

Carrying Out an Empirical Project

4. Econometric Analysis

Given your research question and the data available, you have to decide on the *appropriate econometric methods* to use.

● **Some general guidelines**
  - ✓ *OLS* is still the most widely used method and often appropriate
  - ✓ Make sure the *key assumptions* are satisfied in your model
  - ✓ Always check for possible problems of *omitted variables, self-selection, measurement error, and simultaneity*
  - ✓ Carefully choose **functional form specifications** (logs, squares etc.)
  - ✓ Beginners mistake: do not include variables that are listed as numerical values but have no quantitative meaning (e.g., *3-digit occupations*), and transform such variables to dummy variables representing categories

Carrying Out an Empirical Project

- **Some general guidelines**

  - ✓ Handle *ordinal regressors* in a similar way (e.g., job satisfaction), and for ordinal dependent variables, there are ordered logit/probit models

  - ✓ One can also reduce *ordered variables* to binary variables

  - ✓ Think of *secondary complications* such as heteroskedasticity

  - ✓ Specific problems in time series regressions: 1) levels vs. differences, 2) trends and seasonality, 3) unit roots and cointegration

  - ✓ Carry out *misspecification tests* and think about possible biases

  - ✓ Sensitivity analysis: look at variations of your specification/method; hopefully, results do not change in a substantial way

  - ✓ Are there problems with *outliers/influential observations*?

Carrying Out an Empirical Project

- **Specific aspects to think of when using panel data (skipped)**
  - ✓ Key assumptions
    - Random effects: regressors unrelated to individual specific effects
    - Fixed effects: regressors related to individual specific effects
    - The fixed effects assumption is often more convincing
    - *Contemporaneous exogeneity*: idiosyncratic errors are uncorrelated with the explanatory variables of the same time period
    - *Strict exogeneity*: idiosyncratic errors are uncorrelated with the explanatory variables of all time periods (often problematic)
  - ✓ Methods for panel data
    - Pooled OLS: random effects assumption, serial correlation of error terms, needs only contemporaneous exogeneity
    - Random effects estimation: random effects assumption, more efficient than pooled OLS, needs strict exogeneity
    - *Fixed effects estimation*: fixed effects assumption, problem with time invariant regressors, needs strict exogeneity
    - First differencing: similar to fixed effects, good for longer time series

Carrying Out an Empirical Project

● **Data mining/specification searches** **(skipped)**

- ✓ The process of looking for the best model is called *specification search*

- ✓ Often, one starts with a general model and drops insignificant variables

- ✓ If the specification search entails many steps, this is *problematic*

- ✓ Our assumptions actually require that the model is only estimated once

- ✓ If one sequentially estimates a number of models on the same data, the resulting test statistics and p-values cannot be interpreted anymore

- ✓ This (difficult) problem is often ignored in practice

- ✓ One should *keep the number of specification steps to a minimum*

Carrying Out an Empirical Project

5. Writing an empirical paper

A succesful empirical paper *combines* *a careful, convincing data analysis with good explanations and a clear exposition*

Ⅰ. Introduction

- ✓ State basic objectives and explain why the topic is important

- ✓ Literature review: What has been done? How do you add to this?

- ✓ Grab the reader's attention by presenting simple statistics, paradoxical evidence, topical examples, or challenges to common wisdom

- ✓ One may give *a short summary of results* in the introduction

Carrying Out an Empirical Project

Ⅱ. Conceptual (or theoretical) framework

- ✓ *Description of general approach* to answering your research question: you may delevop/use a formal economic model for this

- ✓ For example, setting up a utility maximization model of criminal activity clarifies the factors that matter for explaining criminal activity

- ✓ However, often common economic sense suffices to discuss the main mechanisms and control variables that have to be taken into account

- ✓ As one is in most cases interested in answering a causal question, a convincing discussion of what variables to control for is essential

Carrying Out an Empirical Project

Ⅲ. **Econometric models and estimation methods** （有时，可与第二或第四部分合并）

- ✓ Specify the population model（总体模型） you have in mind

- ✓ Example: Effects of alcohol consumption on college GPA

$$colGPA = \beta_0 + \beta_1 alcohol + \beta_2 hsGPA + \beta_3 SAT + \beta_4 female + u$$

- ✓ Example: Time series model of city-level car thefts

$$thefts_t = \beta_0 + \beta_1 unem_t + \beta_2 unem_{t-1} + \beta_3 cars_t + \beta_4 convrate_t + \beta_5 convrate_{t-1} + u_t$$

- ✓ Describe how you *measure the variables* in your population model

- ✓ Explain your *functional form choices* and discuss *estimation methods*

  - When using OLS: Discuss why exogeneity assumptions hold, and how you deal with heteroskedasticity, serial correlation, and the like

  - When using IV/2SLS: Explain why your instrumental variables fulfill the assumptions: 1) exclusion, 2) exogeneity, 3) partial correlation

  - When using panel methods: Explain what the unobserved individual specific effects stand for, and how they are removed/accounted for

Carrying Out an Empirical Project

IV. Data

- ✓ Carefully describe the data used in your empirical analysis
- ✓ *Name the sources of your data and how they can be obtained*
- ✓ Time series data and *short data sets* may be listed in the appendix
- ✓ If your data is self-collected, include *a copy of the questionnaire*
- ✓ Discuss the units of measurement of the variables of interest
- ✓ Present *summary / descriptive statistics* for the variables used in the analysis
- ✓ *For trending variables*, growth rates or graphs are more appropriate
- ✓ Always state how many *observations* you use for different estimations

Carrying Out an Empirical Project

Ⅴ. Results

- ✓ *Present estimated equations*, <u>or</u>, if there are too many, *present tables*
- ✓ Always include things like R-squared and the number of observations
- ✓ Are your estimated coefficients *statistically significant*?
- ✓ Are they *economically significant*? What is their magnitude?
- ✓ If coefficients do not have the expected signs, this may indicate there is a *specification problem*, for example, omitted variables
- ✓ Relate *differences between the results from different methods* to the differences in the assumptions underlying these methods

Ⅵ. Conclusion

- ✓ Summarize main results and conclusions from them
- ✓ Discuss *caveats* to the conclusions drawn
- ✓ Suggest directions for further research

Carrying Out an Empirical Project

● **Style hints 1**

- ✓ Choose a title *that is exciting and reflects the paper's topic*
- ✓ Papers should be typed and formatted
- ✓ Number equations, graphs, and tables
- ✓ Refer to papers by author and date, for example, White (1980)
- ✓ When you introduce an equation, describe *important variables*
- ✓ In order to <u>focus on a particular variable</u> (e.g., *alcohol*) you may write something like

$$GPA = \beta_0 + \beta_1 alcohol + \boxed{\mathbf{x\delta}} + u$$

Shorthand for several other explanatory variables

- ✓ Presenting results in equation form:

$$\widehat{salary} = \underset{(223.90)}{830.63} + \underset{(.0089)}{.0163}\, sales + \underset{(11.08)}{19.63}\, roe$$

$$n = 209, R^2 = .029$$

State near the first equation that standard errors are *in parentheses*

Carrying Out an Empirical Project

● Style hints 2

TABLE 19.1 OLS Results. Dependent Variable: Participation Rate

Independent Variables	(1)	(2)	(3)
mrate	.156	.239	.218
	(.012)	(.042)	(.342)
$mrate^2$	—	−.087	−.096
		(.043)	(.073)
log(emp)	−.112	−.112	−.098
	(.014)	(.014)	(.111)
$log(emp)^2$	.0057	.0057	.0052
	(.0009)	(.0009)	(.0007)
age	.0060	.0059	.0050
	(.0010)	(.0010)	(.0021)
age^2	−.00007	−.00007	−.00006
	(.00002)	(.00002)	(.00002)
sole	−.0001	.0008	.0006
	(.0058)	(.0058)	(.0061)
constant	1.213	.198	.085
	(.051)	(.052)	(.041)
industry dummies?	no	no	yes
Observations	3,784	3,784	3,784
R- squared	.143	.152	.162

Note: The quantities in parentheses below the estimates are the standard errors.

© Cengage Learning, 2016

Reporting results in tabular form:

Clearly indicate dependent and independent variables.

Limit the number of digits reported after the decimal point.

You may also think of *rescaling your variables* so that coefficients are not too large or too small.