Research Methodology in Finance
MGMG 522 : Session #1

Charn Soranakom, Ph.D.
Lecturer of Finance
College of Management, Mahidol University

Steps in Writing a Research Paper

1. Select a topic
2. Review the relevant literature
3. Develop a hypothesis to be tested
4. Gather the needed data
5. Select the appropriate model and tool
6. Do the data analysis
7. Interpret the results
8. Conclude the paper

*Note: The first three steps do not have to be in that order.

1. Select a Topic

- First, you must be familiar with the topic you select.
- To become familiar with a particular topic, you should read many scholarly (peer-reviewed) papers in academic journals.
- Use online database to find the papers of your interest. Examples are
  - ProQuest (access through our e-learning system)
  - EBSCO (access through our e-learning system)
  - ScienceDirect (access through MU network)
  - JSTOR (access through MU network)
- You should practice your searching technique. For example, you should know what keyword to use to get the results you wanted.
- The keyword should not be too general or too specific.
- Use “AND”, “OR”, “NOT” to limit your search time and the number of results.

1-3

- You should also observe the trend in research from studying recently published articles in the top journals.
- That way you will learn about the trend, i.e., what are the hot topics being studied these days.
- You should not spend your valuable time studying and writing on the “dead” topics.
- These are some of the leading journals in the field of Finance.
  - Journal of Finance (JF)
  - Journal of Business (JB)
  - Journal of Financial and Quantitative Analysis (JFQA)
  - Review of Financial Studies (RFS), and
  - Journal of Financial Economics (JFE)
  - [See, “Finance Journal Rankings”, 2001]

1-4

Other Sources for Journal Articles

- Social Science Research Network (SSRN) has downloadable working papers, forthcoming papers, and published papers at www.ssrn.com
- Journal of Finance gives a free preview of the forthcoming papers at www.afajof.org/journal/forthcoming.asp
- Review of Financial Studies allows you to view abstracts of past and current papers at rfs.oxfordjournals.org
- Charles A. Dice Center has a list of journal websites, which can be found at www.cob.ohio-state.edu/fin/journal/jofsites.htm

1-5

Research Areas in Finance

This is a list of possible topics (not comprehensive).

- Mergers and Acquisitions
- Mutual Fund Performance
- Ownership Structure
- Dividend Policy
- Share Repurchases
- Earnings and Stock Returns
- Capital Structure
- Efficient Markets (using Event Studies)
- Investment Techniques
- Derivatives
- IPO
- Insider Trading
- Portfolio Management
- Behavioral Finance
- Market Microstructure

1-6

2. Literature Review

- In this step, you will try to form a knowledge base.
- This step will require a lot of reading.
- After doing a lot of reading, you should have some ideas about what other authors have done or discovered.
- With your careful reading, you should be able to identify what has not been done or written about in that topic. At this stage, you will come up with your research question.

1-7

3. Develop a Hypothesis

- Once, you identify your research question, you need to translate your research question into a testable hypothesis.
- Some research questions cannot be tested empirically. You might have to work around that problem.
- A hypothesis is something that can be tested with data using a statistical/econometric tool. Examples are
  - Test of a single mean
  - Comparison of the means (2 groups or more)
  - Comparison of the variances (2 groups or more)
  - Test of relationship (to see if there is + or - relationship)
  - Regression analysis and its variants
  - Time series analysis and forecasting
- Examples of some hypotheses that CANNOT be tested.
  - Cause and effect (e.g., A causes B)
  - Some experiments in human (e.g., human behavior)

1-8

4. Gather the Needed Data

- Your hypothesis should tell you about what kind of data (variable) is needed for your analysis.
- Variables can either be qualitative or quantitative.
- Qualitative type of data needs to be converted into number before performing any analysis (e.g., using dummy or binary variable)
- Some variables cannot be measured directly, hence, need to be proxied.
- Types of variables are
  - Nominal scale (e.g., gender or marital status)
  - Ordinal scale (e.g., grade or education)
  - Interval scale (e.g., year or temperature)
  - Ratio scale (e.g., age or salary)

  Note: These 4 types of data differ on whether the ratio of, the difference between, and the order of two variables are meaningful or not .

1-9

5. Select Your Tool

- Almost all of your analysis will be performed using a statistical or econometric software package.
  - SPSS, Minitab, EViews, Limdep, and SAS
- You should be familiar with the software package you use.
- Generally, the software package will not make a mistake. It is usually the person who makes the mistake.
- Each software package is usually different from the others. If you don't understand something, you should consult the manual. There is a danger in using a tool that you don't know how it operates.
- If you input bad data or input erroneous data, you will not get the correct result. Remember, GIGO (garbage in, garbage out).

1-10

6. Do the Analysis

- When you have a big data set or many data files, you should have or develop a system to keep track of your data files.
- Remember to back up your data.
- You should make printouts (hard copy) periodically. Better be safe than sorry). In case you lost your files due to power outage, data corruption, or virus, you still can recover your data in the unlikely event that your backup files failed.
- Keep track of any changes you have made. Dating or numbering your documents.
- Don't count on your memory. Write instead.

1-11

7. Interpret the Results

- The analysis you have performed will need to be converted into words that explain what you've discovered.
- This part requires much of your communication skills. A good paper explains the results very well to the audience.
- Be clear, concise, and complete.
- Don't assume that the audience is sophisticated and, hence, will understand or be able to read your mind.

1-12

8. Conclude the Paper

- After your hard work, you have to be able to tell the audience about your research in overall in writing.
- In your conclusion, you should summarize your paper in one or two pages.
- The conclusion section is the part where interested readers will learn more (than from reading your abstract) about your research in a short time. The conclusion should state
  - Your research hypothesis
  - Your data and tools for the analysis
  - Your major findings and what they mean

1-13

Writing a Paper

- Writing should not occur all at once after the research is completed.
- Rather, you should write down or update your paper periodically, little by little.
- Writing is an art that requires good communication skills.
- Remember, you will not be there by the reader's side to explain your paper to him/her. The words you write that make up your paper will do the explanation.

1-14

Types of Research Papers

1. Theoretical Paper
   - This type of paper is concerned with developing a model to explain economic or financial phenomena. It usually requires a set of assumptions and is involved with mathematical models. Some of these models can be tested empirically, while some cannot.
2. Empirical Paper
   - This type of paper requires data for the analysis. The relationships and explanations have already been developed or known in previous papers. Your job is to gather the data and test such relationships.

1-15

General Format of an Empirical Paper

1. Introduction and Literature review
   - Tell the author's motivation for doing this research.
   - Review the literature (what have been done or found).
   - Tell how this paper is different from the others.
2. Hypothesis development
   - This part will usually refer to or restate the previous findings.
   - Tells what is to be tested in this paper and what theory backs up the hypothesis.
3. Data and Analysis
   - Tell where the data came from and the tools for data analysis.
4. Results
   - State your results or what you find in your analysis.
   - This part should answer your hypothesis stated earlier in the "Hypothesis Development" section of your paper.
5. Conclusion
   - This is a very important part of your paper.
   - It should briefly summarize your whole paper.
   - It should describe your paper in greater detail than your abstract.
   - Some authors also provide
     - a direction for future research
     - a word of cautions or limitations

1-16

General Format of a Theoretical Paper

1. Introduction and Literature review
   - Similar to that of an empirical paper.
2. Assumptions of the model
   - There are many factors that influence or explain economic behaviors.
   - We need to discard things that are not important and focus only on the thing under the study.
   - Therefore, we need to provide some assumptions to exclude the things we don't consider.
3. The Model
   - Explain economic behaviors with relationships, mathematical formulas, or theory.
4. Results
   - State your results.
5. Conclusion
   - similar to that of an empirical paper.

1-17

Types of Data

- Primary data
  - Data you collect by yourself from the interviews or observations.
- Secondary data
  - Data that already exist in print or electronic format.
  - Some are freely available, some are available from data vendors and require subscription or per-usage fees.
  - Examples of data in electronic format
    - CompuStat (from S&P)
    - DataStream
    - CRSP (from Center for Research in Securities Prices, University of Chicago)
    - SET SMART (from Stock Exchange of Thailand)
  - Examples of data in print format
    - SET publications
    - Newspaper
    - Magazines
    - Company annual reports to shareholders
- Data can also be categorized as time-series or cross-sectional.

1-18

SET SMART

- A web-based database
- There are two versions
  - Investor version (available for purchase): Non real-time data, plus limited historical data
  - Enterprise version: Only historical data dated back to 1975
- Real-time price data are available free of charge at www.set.or.th
- Enterprise version can only be accessed from our campus at http://set.oz.cmmu.net (may require initial set-up)

1-19

Other Considerations

- Plagiarism
- Referencing system and format
- Writing a thesis or independent study
  - Usually there are five chapters in a thesis or independent study. (The names of the chapters can differ from the names given below.)
    - Chapter 1: Introduction
    - Chapter 2: Literature Review
    - Chapter 3: Data and Methodology
    - Chapter 4: Results of the Study
    - Chapter 5: Conclusion
- Publication considerations

1-20

An Overview of Regression Analysis & Ordinary Least Squares (OLS)

(Ch. 1 & 2)

What is Econometrics?

- Econometrics tries to model and quantify actual economic phenomena.
- Economics provides us with the theory that is abstract in nature.
- Econometrics uses real data and quantifies the theory.

Uses of Econometrics

1. To model economic phenomena
   - What X's explain Y?
   - How X's affect Y and what are their magnitudes?
   - For example,
     - Economic theory: Demand = f(Price, Income)
     - Econometrics: Demand = 100-0.5*Price+5*Income
2. To test hypotheses about economic theory
   - How strong is the relationship between each X and Y?
   - Is it significant?
   - Is it what we expected?
   - Is the model a good fit overall?
3. To forecast the future
   - What is the value of Y, given X's?

1-23

Econometric Approaches

- Given a data set to work with, we could use many different approaches to come up with a model or equation.
- One of many econometric approaches is regression analysis.
- By far, regression analysis is the most widely used approach.

1-24

What is Regression Analysis?

- Regression analysis is a statistical method that tries to specify a single equation that describes what, how, and how much X's affect Y.
- Y is called a dependent variable.
- X's are called independent variables.
- Although Y is called a dependent variable, it doesn't mean that Y is dependent on X's, or X's cause Y.
- Regression analysis does not imply cause and effect or causality.

1-25

Single-Equation Linear Models

- A single-equation linear model takes the form,
  $Y = \beta_0 + \beta_1 X$
- It is called "single-equation" because there are no other equations that specify relationships between X and Y.
- It is called "linear" because X is linearly related to Y. In other words, if you were to plot a graph between X and Y, you would get a straight line.
- β_0 is called the intercept. It tells the value of Y if X is zero. We normally do not care much about β_0.
- Our interest is mainly on β_1, which is called the coefficient. It measures the direction and the magnitude of change in Y if X changes by one unit.

1-26

What does “linear” mean?

- A regression equation can be linear in coefficients or linear in variables.
- Linear in coefficients: all β’s must be in their simplest form (sometimes, after a transformation).
- Linear in variables: all X’s must be in their simplest form (sometimes, after a modification). A plot of X against Y must be a straight line.
- Q: Which of the followings are linear in coefficients and which are linear in variables?

  $Y = \beta_0 + \beta_1 X$

  $Y = \beta_0 + \beta_1 \ln X$

  $\ln Y = \beta_0 + \beta_1 X$

  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

  $Y = e^{\beta_0} X_1^{\beta_1} X_2^{\beta_2}$

1-27

What do we mean by “linear” regression?

- We mean that the regression equation must be “linear in coefficients”.
- The regression equation DOES NOT have to be “linear in variables.”
- Therefore, our regression equation can be, $\ln Y = \beta_0 + \beta_1 X^3$, and still satisfies the linearity in coefficients condition.
- Linearity in coefficients condition will always be satisfied if you can write the regression equation in the form,

  $f(Y) = \beta_0 + \beta_1 f(X)$.

1-28

Stochastic or Random Error Term

- Usually the relationship between X and Y is not exact.
- So, there will usually be some variation in Y that cannot be explained completely by X.
- Causes of this random error are
  - Omitted variables,
  - Measurement error,
  - Incorrect functional form, or
  - Purely random error.
- Regression analysis only allows an additive random error term like, $Y = \beta_0 + \beta_1 X + \varepsilon$.
- Regression analysis does not allow a multiplicative random error term like, $Y = \beta_0 + \beta_1 X + X\varepsilon$.

1-29

About a Regression Equation

- A typical regression equation is written as,
  $Y = \beta_0 + \beta_1 X + \varepsilon$
- $\beta_0 + \beta_1 X$ is the deterministic term.
- ε is the stochastic or random error term.
- $Y = \beta_0 + \beta_1 X + \varepsilon$ is the "population" regression equation, which is unknown.
- Given a value of X, we can estimate only the <u>expected</u> value of Y, not the <u>true</u> value of Y.
- Our job is to estimate the "population" regression equation with a sample set of data. As a result, we will get the "estimated" regression equation like, $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

1-30

Estimate a Regression Equation

$Y_1 =$	β_0	$+ \beta_1 * X_1$	$+ \varepsilon_1$
$Y_2 =$	β_0	$+ \beta_1 * X_2$	$+ \varepsilon_2$
$Y_3 =$	β_0	$+ \beta_1 * X_3$	$+ \varepsilon_3$
$Y_4 =$	β_0	$+ \beta_1 * X_4$	$+ \varepsilon_4$
$Y_5 =$	β_0	$+ \beta_1 * X_5$	$+ \varepsilon_5$
...	...	...	...
...	...	...	...
$Y_n =$	β_0	$+ \beta_1 * X_n$	$+ \varepsilon_n$

1-31

Estimation

True Regression Equation	Estimated Regression Equation
β_0	$\hat{\beta}_0 \text{ or } b_0$
β_1	$\hat{\beta}_1 \text{ or } b_1$
ε	e
$Y = \beta_0 + \beta_1 X + \varepsilon$	$Y = \hat{\beta}_0 + \hat{\beta}_1 X + e, \text{or}$ $Y = b_0 + b_1 + e$

1-32

Estimation (continue…)

- ε is not observable.
- But we can observe e (called residuals).

$$e_i = Y_i - \hat{Y}_i$$

$$e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

- There are many ways to come up with the estimates for β_0 and β_1.
- Given data on X_i and Y_i.
- If we change β_0 and/or β_1, e_i will change as well.

1-33

Ordinary Least Squares: OLS

- OLS is a regression estimation method that minimizes the sum of the squared residuals.
- OLS is sometimes called an estimator.
- Estimator = Estimation Method.

$$OLS \text{ minimizes } \sum_{i=1}^{n} e_i^2$$

$$OLS \text{ minimizes } \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

1-34

Why use OLS?

1. OLS is quite easy to use.
2. Minimizing the sum of the squared residuals is a theoretically sound objective.

   * The other objective could be to minimize Σe_i or $\Sigma|e_i|$, but there are problems with this objective.
3. OLS estimates possess a number of useful properties.
   - The regression line passes through the means of X & Y.
   - The sum of the residuals is exactly zero.
   - Under some restrictions, OLS is the best estimator.

1-35

How do we estimate β_0 and β_1?

- The objective function is, $\text{Min} \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$
- To find β_0 and β_1 that minimize the objective function, we take a partial derivative of the objective function with respect to each parameter.

$$\frac{\partial \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-1) = 0$$

$$\frac{\partial \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-X_i) = 0$$

1-36

Normal Equations

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1\left(\sum_{i=1}^{n} X_i\right)$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{\beta}_0\left(\sum_{i=1}^{n} X_i\right) + \hat{\beta}_1\left(\sum_{i=1}^{n} X_i^2\right)$$

Estimates

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left[X_i - \overline{X}\right]\cdot\left[Y_i - \overline{Y}\right]}{\sum_{i=1}^{n}\left[X_i - \overline{X}\right]^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$

TSS, ESS, and RSS

- Total Sum of Squares, $TSS = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$
- Explained Sum of Squares, $ESS = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$
- Residual Sum of Squares, $RSS = \sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2$

- For a given data set, TSS cannot change and does not depend on the estimation method.
- OLS is the method that guarantees to find the estimates of β_0 and β_1 that minimize RSS.
- We usually delegate the task of finding the estimates of β_0 and β_1 to the computer, especially for a multiple regression analysis.

1-39

Coefficients in a Multiple Regression

- The coefficient in a multiple regression measures how much Y will change if the dependent variable in question changes by one unit, holding constant the other independent variables included in the regression equation.
- This is equivalent to a partial derivation of Y with respect to each X.
- Multiple regression analysis technically allows us to hold constant the influences of other variables and focus on the influence of just one variable at a time (where otherwise impossible).

1-40

Goodness of Fit Test: R^2

- We want the regression equation to fit the data well.
- One way to measure the overall goodness of fit is to compare ESS to TSS.
- That measure is R^2.
- R^2 = ESS/TSS = 1 – (RSS/TSS)
- $0 \leq R^2 \leq 1$

1-41

- In a time-series type of data, R^2 of 0.9 may be considered a good fit.
- On the other hand, in a cross-sectional type of data, R^2 of 0.5 may be considered a good fit.
- How high the value of R^2 is considered a good fit is very subjective.

1-42

Adj-R^2

- When you add one more variable to the regression equation, your R^2 value will either stay the same or increase. There is no way that the R^2 value will decline. Why?
- Each time we add one independent variable to the regression equation, we lose one more degree of freedom.
- To adjust for the loss of degrees of freedom as a result of the addition of independent variables, we calculate the adjusted R^2, or Adj-R^2.

$$Adj - R^2 = 1 - \left(1 - R^2\right) \cdot \frac{(n-1)}{(n-K-1)}$$

n = # of observations, K = # of independent variables.

1-43

Cautions about R^2 and Adj-R^2

- Both R^2 and Adj-R^2 measure the overall goodness of fit only.
- Obtaining a high value of R^2 or Adj-R^2 for the regression equation is desirable, but it is not all that matters!
- Comparing two competing regression models on the basis of R^2 or Adj-R^2 can be misleading and dangerous.
- You should evaluate the competing equations with other criteria (which will be discussed in later session).

1-44