SBNM 5220
ECONOMETRICS 2 sh
READING GUIDE
Spring, Quad B

Instructor: Lee Sundholm
5043 N. Spaulding Ave.
Chicago, Illinois 60625
Office Phone: 773 – 244 - 5715
Office Fax: 773 - 244 - 5285

e-mail lsundholm@northpark.edu

SBNM 5220
ECONOMETRICS
READING GUIDE

Text: Damodar N. Gujarati, *Essentials of Econometrics*, Third edition, McGraw-Hill, Boston, 2006. **ISBN:** 0072970928 (**later editions are available**)

Course Description

This course serves as an introduction to econometric methods. Econometric methods are used in comparing and testing the validity of alternative theories, as well as assisting in business and economic forecasting. These methods enable students to understand the functioning of the economy and business system, and equip them with an enhanced ability to predict the impact of change.

Econometrics brings together basic mathematical and statistical tools, along with economics and business concepts which enhance the problem solving capabilities of students. Students will learn to apply mathematical methods to contemporary business and economic problems and issues. Mathematical methods are applied in both macro and micro environments in order for students to gain broad knowledge of the quantitative approach to problem solving.

Students will come to understand how quantitative methods provide a rigorous approach to resolving issues of choice in the implementation alternative policies. Students will learn how the application of econometric methods adds to the reliability of analysis required for appropriate economic, business policy, and strategic managerial decisions.

Course Objectives

This course serves as an introduction to econometric methods. Econometric methods are used in comparing and testing the validity of alternative theories, as well as assisting in business and economic forecasting. These methods enable students to understand the functioning of the economy and business system, and equip them with an enhanced ability to predict the impact of change.
Econometrics brings together basic mathematical and statistical tools, along with economics and business concepts which enhance the problem solving capabilities of students. Students will learn to apply mathematical methods to contemporary business and economic problems and issues. Mathematical methods are applied in both macro and micro environments in order for students to gain broad knowledge of the quantitative approach to problem solving.
Students will come to understand how quantitative methods provide a rigorous approach to resolving issues of choice in the implementation alternative policies. Students will learn how the application of econometric methods adds to the reliability of analysis required for appropriate business policy and strategic decisions.

Course Grading

Three examinations will be given during the course. A course paper on an approved research project will be required. Each of these four requirements will be weighted at 25%. The course paper is due two weekdays after the last class.

Academic Honesty

In keeping with our Christian heritage and commitment, North Park University and the School of Business and Nonprofit Management are committed to the highest possible ethical and moral standards. Just as we will constantly strive to live up to these standards, we expect our students to do the same. To that end, cheating of any sort will not be tolerated. Students who are discovered cheating are subject to discipline up to and including failure of a course and expulsion.

Our definition of cheating includes but is not limited to:

1. Plagiarism—the use of another's work as one's own without giving credit to the individual. This includes using materials from the internet.

2. Copying another's answers on an examination.

3. Deliberately allowing another to copy one's answers or work.

4. Signing an attendance roster for another who is not present.

In the special instance of group work, the instructor will make clear his/her expectations with respect to individual vs. collaborative work. A violation of these expectations may be considered cheating as well.

For further information on this subject you may refer to the Academic Dishonesty section of the University's online catalog.

In conclusion, it is our mission to prepare each student for a life of significance and service. Honesty and ethical behavior are the foundation upon which such lives are built. We therefore expect the highest standards of each student in this regard.

Students with Disabilities

Students with a disabilities who believe that they may need accommodations in this class are encouraged to contact your program's office (Education: 773-244-5730; Business: 773-244-6270; Nursing: 773-244-5680; Community Development: 773-244-5774; Seminary: 773-244-6219; Music: 773-244-5625). Please do so as soon as possible to better ensure that such accommodations are implemented in a timely manner. For further information please review the following website: http://www.northpark.edu/campus-Life-and-Services/disability-resources

Class Attendance Policy

The graduate courses in the SBNM are all 7 weeks in length. Missing one class session is allowed without penalty as long as all readings and assignments are made up by the student within a reasonable time period (the following class session). Missing a second class session is allowed only in unusual circumstances by prior arrangement with the instructor. Since this represents 25% of the class time for the course, the student runs the risk of receiving a lower overall grade for the class. A student who misses three classes will automatically fail the course, unless the student drops the course before the eighth week of class. Students who drop a course will be held responsible for tuition, based upon the current North Park University policy outlined in the University Catalog (found on the NPU website).

Use of *APA Publication Manual*

The School of Business and Nonprofit Management (SBNM) has adopted the *Publication Manual of the American Psychological Association* (APA) as the standard and required format for all written assignments in SBNM courses.

Our goal in adopting the *APA Manual* is to enhance student learning by:

1) Improving student's writing skills.
2) Standardizing the required format of all written assignments in all SBNM courses.
3) Emphasizing the importance of paper mechanics, grammatical constructs, and the necessity of proper citations.
4) Holding students accountable for high quality written work.

If you are unfamiliar with the requirements of the *APA Publication Manual*, we recommend that you purchase the reference manual and/or that you consult one or more of the suggested resources as listed on the Student Resources section of the SBNM website. ***It is your responsibility to learn and ensure that all written work is formatted according to the standards of the APA Manual.***

ASSIGNMENTS

Week Chapters in Damodar N. Gujarati, *Essentials of Econometrics*

1 Chapters 1, 2

DESCRIPTION OF MATERIAL

Research in economics, finance, management, marketing, and related disciplines is becoming increasingly quantitative. This section provides an overview of what econometrics is about, and begins with a study of mathematical methods.

In econometrics, mathematical statistics plays an important role, and the foundation of mathematical statistics is based on probability theory. The fundamental concepts of probability, includes sample space, sample points, events, random variables, and probability distributions of random variables. In econometrics we deal with relationships between economic and business variables, and deal with the joint probability distributions of such variables. We will discuss joint events and joint variables and their probability distributions, conditional probability distributions, unconditional probability distributions, and statistical independence. An application of the conditional probability distribution is Bayes' Theorem which shows how experimental knowledge can be used to revise probabilities.

Topics include

- Course purposes and overview
- Econometric models
- Review of data sources on the web: text, pp. 16 - 18
- Probability and statistics
- Problem assignments

2 Chapters 3, 4, 5

DESCRIPTION OF MATERIAL

The major characteristics or moments of probability distributions of random variables, such as the expected value, variance, covariance, correlation, skewness, kurtosis, conditional expectation and conditional variance will be introduced. In addition, formulas will tell us how to compute the probabilities of random variables and how to estimate the characteristics of probability distributions (i.e., the moments), such as the expected or mean value, variance, covariance, correlation, and conditional expectation.

Probability distributions have been discussed. Next we consider four probability distributions--the normal, the *t*, the chi-square, and the *F*--and the special features of each distribution. These four PDFs play a very pivotal role in econometric theory and practice. In addition, estimating population parameters on the basis of sample information and testing hypotheses about them are the two main branches of (classical) statistical inference.

Topics include

- Applications to economics and business
- Estimation and hypothesis testing
- Statistical inference
- Problem assignments

3	EXAMINATION 1
4 - 5	Chapters 6, 7, 8

DESCRIPTION OF MATERIAL

We begin work with regression analysis. Starting with the population regression function (PRF), the concept of linear PRF is introduced. Regressions are *linear in the parameters* regardless of whether or not they are linear in the variables. The idea of the stochastic PRF is presented along with the stochastic error term *u*. PRF is a theoretical or idealized construct because, in practice, all we have is a sample(s) from some population. Then the sample regression function (SRF) is introduced. In order to answer the question of how the SRF is obtained we consider the method of ordinary least squares (OLS), and present appropriate formulas to estimate the parameters of the PRF. Next we determine how good the SRF obtained by OLS is as an estimator of the true PRF. We will show how the estimated model can be used for the purpose of drawing inferences about the true population regression model. Although the two-variable model is the simplest possible linear regression model, the ideas introduced are the foundation of the more involved multiple regression models.

The simplest of the multiple regression models is the three-variable linear regression model--one dependent variable and two explanatory variables. The three-variable model introduces some new concepts, such as partial regression coefficients, adjusted and unadjusted multiple coefficient of determination, and multicollinearity.

The estimation of the parameters of the multiple regression coefficients is found within the framework of the classical linear regression model using the method of ordinary least squares (OLS). The OLS estimators of the multiple regression

possess several desirable statistical properties summed up in the Gauss-Markov property of best linear unbiasedness (BLUE).

With the assumption that the disturbance term follows the normal distribution with zero mean and constant variance σ^2, each estimated coefficient in the multiple regression follows the normal distribution with a mean equal to the true population value and the variances given by the formulas (developed in the text). In practice, σ^2 is not known and has to be estimated. The OLS estimator of σ^2 is $\hat{o}^2$ the unknown variance. If we use this replacement then, as in the two-variable case, each estimated coefficient of the multiple regression follows the *t* distribution, not the normal distribution.

The knowledge that each multiple regression coefficient follows the *t* distribution with d.f. equal to (*n* - *k*), where *k* is the number of parameters estimated (including the intercept), means we can use the *t* distribution to test statistical hypotheses about each multiple regression coefficient individually. This can be done on the basis of either the *t* test of significance or the confidence interval based on the *t* distribution. In this respect, the multiple regression model does not differ much from the two-variable model, except that proper allowance must be made for the d.f., which now depend on the number of parameters estimated.

When testing the hypothesis that all partial slope coefficients are simultaneously equal to zero, the individual *t* testing referred to earlier is of no help. Here we should use the analysis of variance (ANOVA) technique and the attendant *F* test. Incidentally, testing that all partial slope coefficients are simultaneously equal to zero is the same as testing that the multiple coefficient of determination R^2 is equal to zero. Therefore, the *F* test can also be used to test this latter but equivalent hypothesis. We determine when to add a variable or a group of variables to a model, using either the *t* test or the *F* test. In this context we also discuss the method of restricted least squares.

Topics include

- Linear regression
- Applications to economics and business
- Further regression models: multiple regression, functional models, log models
- ANOVA
- Problem assignments

ASSIGNMENTS

Week **Chapters in Damodar N. Gujarati, *Essentials of Econometrics***

6 – 7 9, 10, 11, 12

DESCRIPTION OF MATERIAL

We discuss models that are linear in parameters, or that can be rendered as such with suitable transformation, but that are not necessarily linear in variables. There are a variety of such models, each having special applications. We consider five major types of nonlinear-in-variable but linear-in-parameter models:

1. The log-linear model, in which both the dependent variable and the explanatory variable are in logarithmic form.
2. The log-lin or growth model, in which the dependent variable is logarithmic but the independent variable is linear.
3. The lin-log model, in which the dependent variable is linear but the independent variable is logarithmic.
4. The reciprocal model, in which the dependent variable is linear but the independent variable is not.
5. The polynominal model, in which the independent variable enters with various powers.

Of course, there is nothing that prevents us from combining the features of one or more of these models. Thus, we can have a multiple regression model in which the dependent variable is in log form and some of the *X* variables are also in log form, but some are in linear form.

We studied the properties of these various models in terms of their relevance in applied research, their slope coefficients, and their elasticity coefficients. We also showed with several examples the situations in which the various models could be used. Needless to say, we will come across several more examples in the remainder of the text.

In choosing among the competing models, the overriding objective should be the economic relevance of the various models and not merely the summary statistics, such as R^2.

- **Model building** *requires a proper balance of theory, availability of the appropriate data, a good understanding of the statistical properties of the various models, and practical judgment.* Since the theory underlying a topic of interest is not perfect, there is no such thing as a perfect model. We attempt to develop a reasonably good model that will balance all these criteria.

Whatever model is chosen in practice, we have to pay careful attention to the units in which the dependent and independent variables are expressed, for the interpretation of the regression coefficients may hinge upon units of measurement.

We will discuss how qualitative, or dummy, variables taking values of 1 and 0 can be introduced into regression models alongside quantitative variables. The dummy variables are essentially a data-classifying device in that they divide a sample into various subgroups based on qualities, or attributes (gender, marital status, race, religion, for example), and *implicitly* run individual regressions for each subgroup. Now if there are differences in the responses of the dependent variable to the variation in the quantitative variables in the various subgroups, they will be reflected in the differences in the intercepts or slope coefficients of the various subgroups, or both. Although it is a versatile tool, the dummy variable technique has to be handled carefully. We should observe the following rules:

1. If the regression model contains a constant term (as most models do), the number of dummy variables must be one less than the number of classifications of each qualitative variable.

2. The coefficient attached to the dummy variables must always be interpreted in relation to the control, or benchmark, group—the group that gets the value of zero.

3. If a model has several qualitative variables with several classes, introduction of dummy variables can consume a large number of degrees of freedom (d.f.). Therefore, we should weigh the number of dummy variables to be introduced into the model against the total number of observations in the sample.

We also discuss the possibility of committing a *specification error,* that is, of fitting the wrong model to the data. If intercepts as well as slopes are expected to differ among groups, we should build a model that incorporates both the *differential* intercept and slope dummies. In this case a model that introduces only the differential intercepts is likely to lead to a specification error. Of course, it is not always easy a priori to find out which is the true model. Thus, some amount of experimentation is required in a concrete study, especially in situations where theory does not provide much guidance.

The linear probability model (LPM) in which the dependent variable is itself binary is introduced. Although LPM can be estimated by ordinary least square (OLS), there are several problems with a routine application of OLS. Some of the problems can be resolved easily and some cannot. Therefore, alternative estimating procedures are needed. Additional concepts are discussed as follows:

1. The classical linear regression model assumes that the model used in empirical analysis is "correctly specified."
2. The term *correct specification* of a model can mean several things, including:
    a. No theoretically relevant variable has been excluded from the model.
    b. No unnecessary or irrelevant variables are included in the model.
    c. The functional form of the model is correct.
    d. There are no errors of measurement.
3. If a theoretically relevant variable(s) has been excluded from the model, the coefficients of the variables retained in the model are generally biased as well as inconsistent, and the error variance and the standard errors of the OLS estimators are biased. As a result, the conventional *t* and *F* tests remain of questionable value.
4. Similar consequences ensue if we use the wrong functional form.
5. The consequences of including irrelevant variables(s) in the model are less serious in that estimated coefficients still remain unbiased and consistent, the error variance and standard errors of the estimators are correctly estimated, and the conventional hypothesis-testing procedure is still valid. The major penalty we pay is that estimated standard errors tend to be relatively large, which means parameters of the model are estimated rather imprecisely. As a result, confidence intervals tend to be somewhat wider.
6. In view of the potential seriousness of specification errors, in this chapter we considered several diagnostic tools to help us find out if we have the specification error problem in any concrete situation. These tools include a graphical examination of the residuals and more formal tests, such as MWD and RESET.

An important assumption of the classical linear regression model is that there is no exact linear relationship(s), or multicollinearity, among explanatory variables. Although cases of exact multicollinearity are rare in practice, situations of near exact or high multicollinearity occur frequently. In practice, therefore, the term *multicollinearity* refers to situations where two or more variables can be highly linearly related.

The consequences of multicollinearity are as follows:

1. In cases of perfect multicollinearity we cannot estimate the individual regression coefficients or their standard errors.
2. In cases of high multicollinearity individual regression coefficients can be estimated and the OLS estimators retain their BLUE property. But the standard errors of one or more coefficients tend to be large in relation to their coefficient values, thereby reducing *t* values. As a result, based on estimated *t* values, we can say that the coefficient with the low *t* value is not statistically different from zero. In other words, we cannot assess the *marginal* or *individual* contribution of the variable whose *t* value is low. Recall that in a multiple regression the slope coefficient of an *X* variable is the *partial regression coefficient,* which measures the (marginal or individual) effect of that variable on the dependent variable, holding all other *X* variables constant. However, if the objective of study is to estimate a group of coefficients fairly accurately, this can be done so long as collinearity is not perfect.

Topics include:

- Applications to economics and business
- Use of dummy variables
- Model building
- OLS
- LPM
- Regression analysis
- Multicollinearity
- Double log and log linear models
- Cobb-Douglas production function
- Returns to scale
- Linear trend model
- Polynomial regression models
- Problem assignments