Applied Econometrics

Professor V.M. Auken, PhD

Applied Econometrics

- The course will cover several key models as well as identification and estimation methods used in modern econometrics.
- We shall begin with exploring some leading models of econometrics, then seeing structures, then providing methods of identification, estimation, and inference.
- You will learn the modern ways of setting up problems and doing better estimation and inference than the current empirical practice.
- You will learn generalized method of moments, the method of M-estimators, as well more modern versions of these methods dealing with important issues such as weak identification or biases arising in high dimensions.
- At the end of the course, we shall also explore very high dimensional formulations, or "big data", where some penalization via lasso or ridge methods is necessary to say anything useful. You will get a lot of hands-on experience with using the methods on real data sets.

Applied Econometrics

- **Textbooks**

Jack Johnston, John DiNardo. *Econometrics Methods*. 4th ed. McGraw-Hill: 1997. ISBN: 0-07-913121-2.

Ахтямов А.М. *Математические модели экономических процессов*. Уфа: РИЦ БашГУ, 2009. ISBN: 978-5-9221-0994-9.

Applied Econometrics

- **Brief Course Outline**

Economic and econometrics models. Ordinary least squares. The classical linear model.

Maximum likelihood estimation (MLE). Consistency of MLE. The score function. Asymptotic normality of MLE. The Cramer-Rao lower bound.

Asymptotic properties of the least squares estimator. Asymptotic normality. Asymptotic efficiency.

Restrictions and hypothesis tests. Exact linear restrictions. Testing.

Generalized least squares (GLS).The GLS estimator. Feasible GLS. Heteroscedasticity.

Stochastic regressors. Cases.

Data problems. Collinearity. Measurement error. Missing observations.

Functional form and nonnested tests. Flexible functional forms. Testing nonnested hypothesis.

Applied Econometrics

- **Brief Course Outline**

Exogeneity and simultaneity.

Limited dependent variables. Choice between two objects: the probit model. The Newton method.

Models for time series data. ARMA models.

Notation and review. Convergenge modes. Rates of convergence and asymptotic equality.

Asymptotic properties of extremum estimators.

Numerical optimization methods. Derivative-based methods. Simulated Annealing.

Generalized method of moments (GMM).

Regression Analysis

Regression analysis includes the following steps:

- Statement of the problem under consideration
- Choice of relevant variables
- Collection of data on relevant variables
- Specification of model
- Choice of method for fitting the data
- Fitting of model
- Model validation and criticism
- Using the chosen model(s) for the solution of the posed problem and forecasting

1. Statement of the problem under consideration

The first important step in conducting any regression analysis is to specify the problem and the objectives to be addressed by the regression analysis.

The wrong formulation or the wrong understanding of the problem will give the wrong statistical inferences.

The choice of variables depends upon the objectives of study and understanding of the problem.

For example, height and weight of children are related. Now there can be two issues to be addressed.

1. Statement of the problem under consideration

(i) Determination of height for given weight, or

(ii) determination of weight for given height

In the case (i), the height is response variable whereas weight is response variable in case (ii).

The role of explanatory variables are also interchanged in the cases (i) and (ii).

1. Statement of the problem under consideration

(i) Determination of height for given weight, or

(ii) determination of weight for given height

In the case (i), the height is response variable whereas weight is response variable in case (ii). The role of explanatory variables are also interchanged in the cases (i) and (ii).

2. Choice of potentially relevant variables

Once the problem is carefully formulated and objectives have been decided, the next question is to choose the relevant variables. It has to kept in mind that the correct choice of variables will determine the statistical inferences correctly.

For example, in any agricultural experiment, the yield depends on explanatory variables like quantity of fertilizer, rainfall, irrigation, temperature etc.

These variables are denoted by $X_1, X_2, \ldots, X_k$ as a set of k explanatory variables.

3. Collection of data on relevant variables

Once the objective of study is clearly stated and the variables are chosen, the next question arises is to collect data on such relevant variables.

The data is essentially the measurement on these variables.

For example, suppose we want to collect the data on age. For this, it is important to know how to record the data on age. Then either the date of birth can be recorded which will provide the exact age on any specific date or the age in terms of completed years as on specific date can be recorded.

3. Collection of data on relevant variables

Moreover, it is also important to decide that whether the data has to be collected on variables as quantitative variables or qualitative variables.

For example, if the ages (in years) are 15, 17, 19, 21, 23, then these are quantitative values. If the ages are defined by a variable that takes value 1 if ages are less than 18 years and 0 if the ages are more than 18 years, then the earlier recorded data is converted to 1, 1, 0, 0, 0.

Note that there is a loss of information in converting the quantitative data into qualitative data.

3. Collection of data on relevant variables

The methods and approaches for qualitative and quantitative data are also different.

If the study variable is binary, then logistic and probit regressions etc. are used.

If all explanatory variables are qualitative, then analysis of variance technique is used.

If some explanatory variables are qualitative and others are quantitative, then analysis of covariance technique is used. The techniques of analysis of variance and analysis of covariance are the special cases of regression analysis.

3. Collection of data on relevant variables

Generally, the data is collected on n subjects, then y denotes the response or study variable and $y_1, y_2, \ldots, y_n$ are the n values. If there are k explanatory variables $X_1, X_2, \ldots, X_k$ then x_{ij} denotes the i^{th} value of j^{th} variable, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, k$. The observation can be presented in the following table:

3. Collection of data on relevant variables

Notations for the data used in regression analysis

Observation Number	Response y	Explanatory variables			
		X_1	X_2	$\cdots$	X_k
1	y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
3	y_3	x_{31}	x_{32}	$\cdots$	x_{3k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

4. Specification of model

The experimenter or the person working in the subject usually helps in determining the form of the model. Only the form of the tentative model can be ascertained and it will depend on some unknown parameters. For example, a general form will be like

$$y = f(X_1, X_2, ..., X_k; \beta_1, \beta_2, ..., \beta_k) + \varepsilon$$

where ε is the random error reflecting mainly the difference in the observed value of y and the value of y obtained through the model. The form of $f(X_1, X_2, ..., X_k; \beta_1, \beta_2, ..., \beta_k)$ can be linear as well as nonlinear depending on the form of parameters $\beta_1, \beta_2, ..., \beta_k$. A model is said to be linear if it is linear in parameters. For example,

4. Specification of model

$$y = \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \varepsilon$$

$$y = \beta_1 + \beta_2 \ln X_2 + \varepsilon$$

are linear models whereas

$$y = \beta_1 X_1 + \beta_2^2 X_2 + \beta_3 X_2 + \varepsilon$$

$$y = (\ln \beta_1) X_1 + \beta_2 X_2 + \varepsilon$$

are non-linear models. Many times, the nonlinear models can be converted into linear models through some transformations. So the class of linear models is wider than what it appears initially.

4. Specification of model

If a model contains only one explanatory variable, then it is called as simple regression model.
When there are more than one independent variables, then it is called as multiple regression model.
When there is only one study variable, the regression is termed as univariate regression.
When there are more than one study variables, the regression is termed as multivariate regression.
Note that the simple and multiple regressions are not same as univariate and multivariate regressions.
The simple and multiple regression are determined by the number of explanatory variables whereas univariate and multivariate regressions are determined by the number of study variables.

5. Choice of method for fitting the data

After the model has been defined and the data have been collected, the next task is to estimate the parameters of the model based on the collected data. This is also referred to as parameter estimation or model fitting.

The most commonly used method of estimation is the least squares method. Under certain assumptions, the least squares method produces estimators with desirable properties.

The other estimation methods are the maximum likelihood method, ridge method, principal components method etc.

6. Fitting of model

The estimation of unknown parameters using appropriate method provides the values of the parameters. Substituting these values in the equation gives us a usable model. This is termed as model fitting.

The estimates of parameters $\beta_1, \beta_2, \ldots, \beta_k$ in the model

$$y = f(X_1, X_2, \ldots, X_k, \beta_1, \beta_2, \ldots, \beta_k) + \varepsilon$$

are denoted as $\ddot{\beta}_1, \ddot{\beta}_2, \ldots, \ddot{\beta}_k$. which gives the fitted model as

$$y = f(X_1, X_2, \ldots, X_k, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k).$$

6. Fitting of model

When the value of y is obtained for the given values of $X_1, X_2, ..., X_k$, it is denoted as $\hat{y}$ and called as fitted value.

The fitted equation is used for prediction. In this case, $\hat{y}$ is termed as **predicted value.** Note that the fitted value is where the values used for explanatory variables correspond to one of the *n* observations in the data whereas predicted value is the one obtained for any set of values of explanatory variables. It is not generally recommended to predict the y - values for the set of those values of explanatory variables which lie outside the range of data. When the values of explanatory variables are the future values of explanatory variables, the predicted values are called forecasted values.

7. Model criticism and selection

The validity of statistical method to be used for regression analysis depends on various assumptions.
These assumptions are essentially the assumptions for the model and the data. The quality of statistical inferences heavily depends on whether these assumptions are satisfied or not.

For making these assumptions to be valid and to be satisfied, care is needed from beginning of the experiment. One has to be careful in choosing the required assumptions and to examine whether the assumptions are valid for the given experimental conditions or not. It is also important to decide the situations in which the assumptions may not meet.

7. Model criticism and selection

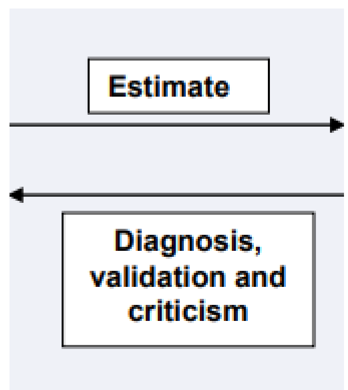
The validation of the assumptions must be made before drawing any statistical conclusion. Any departure from validity of assumptions will be reflected in the statistical inferences.

In fact, the regression analysis is an iterative process where the outputs are used to diagnose, validate, criticize and modify the inputs.

The iterative process is illustrated in the following figure.

7. Model criticism and selection

- Theories
- Model
- Assumptions
- Data
- Statistical methods



- Estimation of parameters
- Confidence regions
- Tests of hypotheses
- Graphical displays

8. Objectives of regression analysis

The determination of explicit form of regression equation is the ultimate objective of regression analysis.
It is finally a good and valid relationship between study variable and explanatory variables.

The regression equation helps in understanding the interrelationships among the variables. Such regression equation can be used for several purposes.
For example, to determine the role of any explanatory variable in the joint relationship in any policy formulation, to forecast the values of response variable for given set of values of explanatory variables.

QUESTIONS!