Categorical Dependent Variables Estimations With Some Empirical Applications

Alhassan Abdul-Wakeel Karakara, Evans S. C. Osabuohien

OnDemand:
(Individual Chapters)
 Available

$37.50

Current Special Offers

Abstract

Microeconomic datasets are usually large, mainly survey data. These data are samples of hundreds of respondents or group of respondents (e.g., households). The distributions of such data are mostly not normal because some responses/variables are discrete. Handling such datasets poses some problems of summarizing/reporting the important features of the data in estimations. This study concentrates on how to handle categorical variables in estimation/reporting based on theoretical and empirical knacks. This study used Ghana Demographic and Health Survey data for 2014 for illustration and elaborates on how to interpret results of binary and multinomial outcome regressions. Comparison is made on the different binary models, and binary logit is found to be weighted over the other binary models. Multinomial logistic model is best handled when the odds of one outcome versus the other outcome are independent of other outcome alternatives as verified by the Independent of Irrelevant Alternatives (IIA). Conclusions and suggestions for handling categorical models are discussed in the study.

Chapter Preview

Top

Introduction

Econometric data analyses deal mostly with finding the 'maxim' of a statement of hypothesis or preposition with a view to finding a causal relationship among variables. Some econometric models are conducted in order to estimate the characteristics of variables. That is, they are performed in order to determine if a change in one variable influences the other. Examples include analysis of supply and price of goods, consumption pattern of individuals or groups, the interest rate in the financial market, and personal savings. Also, econometric analysis develops economic ways of forecasting the behavior and or trend of economic variables to help in decision-making and policy planning processes.

Models used in econometric analysis give a numerical estimated result based on modelling the pattern of variables to help predict phenomenon in a giving situation. The potential uses of econometrics depend on the degree to which a model reflects the objective; availability, nature, and quality of the data; and techniques employed in the evaluation as well as data generating process (Verbeek, 2004). In some instances, econometric analysis makes it possible to use factual material in order to consolidate and verify theoretical hypotheses and models (Greene, 2003). The type of model used depends on the nature of the variables in estimation and the nature of the relationship among the variables (e.g., models may be stochastic or already determined, may be linear or nonlinear, continuous, or discrete).

The kind of econometric model or analysis adopted depends on, among other things; the nature of the data, the data generating process, and objective of the study. Investigating an aggregate issue like economic growth with the use of gross domestic products (GDP), gross investment, aggregate consumption, among other variables, for an extended period could best be analyzed with a time series analysis. Analyzing survey data with information for a group of respondents at a particular point in time could be performed with cross-sectional models. This depends on the nature of the data and the objective of the analysis. Also, studying the same respondents over a longer period at a successive interval of data gathering could be done with micro econometric panel analysis. Again, studying a group of countries or firms over a longer period is mostly done with panel analysis.

Of the aforementioned econometric ways of data analyses, time series is most often straightforward as its data may be highly balanced with no missing observations. Cross country panel data is also without missing data points or observations. Survey data, which is mainly used for cross-sectional analysis, contain problems of missing observations of interest, nonresponse to essential information from respondents, and categorical nature of the dependent variable. For this reason, many scholars find it to be challenging and uncomfortable dealing with survey data analysis. Therefore, this study examines the issue of categorical dependent variable analysis using empirical household survey data for illustration.

The remainder of the study is structured as follows. The second section looks at the methods and applications of categorical dependent variables as well as theoretical explanations. Section three follows with the presentation of empirical analysis of categorical dependent variable using household survey data. Section four concludes and suggests categorical data analysis.

Key Terms in this Chapter

Probit: Just like the logit, probit also translates the values of the independent variables to range '0' to '1' and uses the standard normal distribution function. Thus, the error term is standard normal distribution. Probit has almost similar results as logit.

Linear Probability Model (LPM): LPM s a probability model that allows the independent variables (Xi) to assume negative infinite and positive infinite values. It is because the estimated probabilities lie outside the 0 – 1 bounds. The LPM does away with the discrete nature of the dependent variable, and the error term violates the assumption of normality.

Multinomial Models: Multinomial models have categorical dependent variables that take on more than two outcome possibilities. An example is a choice of cooking fuel type by a household where a household may choose electricity, LPG, charcoal, or wood. The categories should be mutually exclusive for the model to be the best fit.

Categorical Outcome Variable: These are dependent variables that have mutually exclusive outcomes. That is, the choice of one outcome means non-use of the other outcome. An example is a household that may choose to use charcoal, LPG, or electricity for cooking but not the use of two of these categories at a time.

Binary Models: Models that have a categorical dependent variable that take on two outcome possibilities. For example, in a model of having access to a mobile phone or not, there is only two possible outcomes, whether an individual has a mobile phone or not. Binary models include binary logit and binary probit.

Logit: Logit is a categorical dependent variable model (probability model) that translates the values of the independent variables (Xi), which ranges from negative infinity to positive infinite, into a probability to range from '0' to '1' and compel the disturbance term to be homoscedastic and thus becomes logistically distributed.

Complete Chapter List

Search this Book:[   ]
Reset

Table of Contents

Socioeconomic Influences on Fertility Rate Fluctuations in Developed and Developing Economies (pages 141-163)

Kayla M. Good, Anthony M. Maticic Jr.

$37.50
Add to Cart

Chapter 8

Categorical Dependent Variables Estimations With Some Empirical Applications (pages 164-189)

Alhassan Abdul-Wakeel Karakara, Evans S. C. Osabuohien

$37.50
Add to Cart

Chapter 9

Financial Determinants Affecting Leasing Policies: Empirical Evidence From the Airline Industry (pages 190-213)

Kasım Kiracı

$37.50
Add to Cart

Chapter 10

Does Regional Variation in Startup Concentration Predict Employment Growth in Rural Areas of Ohio, Pennsylvania, and West Virginia? (pages 214-242)

Yaya Sissoko, Brian W. Sloboda

$37.50
Add to Cart