Chapter 10
Qualitative Variables in Econometric Models: Outline

William H. Greene

Abstract
This chapter will examine the evolution of an important topic in the theory and practice of modern Microeconometrics. Most econometric modeling analyzes the covariation of measured outcomes such as economic aggregates (GDP, Employment, Inflation) or microeconomic outcomes such as production, cost, efficiency, profits or hours worked. Contemporary microeconometric research involves two additional threads. (1) Discrete outcomes are nonquantitative – they are indicators of the occurrence of economic states, decisions or events. Consider, for example, behavioral models of commuter choice of travel mode, purchasing decisions of insurance, consumer choice of product brand or self-assessed health or wellbeing rated on a Likert scale. (2) Unobservable variables represent ideas, tendencies or concepts of measurements, such as permanent income, aversion to public transportation or strength of preference for environmental stewardship. These types of variables mandate extensions of the familiar modeling techniques. As will emerge below, they are natural companions in an area of Microeconometrics. This chapter will explore the econometric methods that have been developed to accommodate these extensions of conventional regression-style econometrics.

10.1 Introduction

The subject of this chapter is the econometric models that have been developed to accommodate qualitative variables. Models for discrete outcomes and models for latent variables are both foundational segments of the literature, with long histories. The former appeared first in the 1930s in Bioassay and the latter dated from the social science literature of the 1960s (with antecedents in the statistics literature of

William H. Greene
University of South Florida, Tampa, Florida, USA, and New York University, New York, USA, e-mail: wgreene@stern.nyu.edu

the 1930s). The main theme of the book is the evolution of econometric methods and models: why some approaches aged well while others did less so. That aspect of qualitative variables modeling will appear here as well. Accordingly, we will only develop a somewhat superficial treatment of the technical features of the many variants of qualitative variable models that have appeared in the literature. (See, for example, Greene and Hensher (2010) for a more complete survey of ordered choice models and Greene and Hensher (2026) for an extensive survey of hybrid choice models.) Like the other topics in this book, our subject has evolved, and techniques and model forms have come and gone – for example, the linear probability model, the multinomial – multiperiod probit model and the heteroskedastic binary choice model. Considering our topics from a narrower point of view, we note that the probit (binary choice) model is the standard platform for broad instruction about nonlinear econometric modeling. Some of the received econometric theory, not all of which is unanimously agreed upon, originates in discrete choice modeling. For example, what little is known about the incidental parameters problem in fixed effects econometric modeling is anchored in the binary logit model. We will note several cases in passing.

This survey will examine some particular Econometric models for qualitative variables. It is not exhaustive with respect to Microeconometrics in general or the literature on qualitative variables in particular. Rather, we will survey several of the important issues in the field, including broad topics in econometrics, such as the *incidental parameters problem* and narrow, model specific topics such as interpreting coefficients in the multinomial logit model. The chapter is arranged as follows: Section 10.2 identifies the notion of qualitative variables in an econometric model. We present a study from the received literature that we will use and extend to build a running example. Section 10.3 notes some specific features and implications of qualitative variables. Section 10.4 focuses on the tentpole of modeling with discrete variables, the binary response model (mainly the probit model (Bliss (1934)). This section presents some fundamental results for this style of modeling. Section 10.5 details the two multinomial discrete choice models, ordered and unordered, that account for most of the remaining applications. Section 10.6 introduces models that involve continuous unobservable variables. This is an enormous literature. In this review, we will focus on the narrow segment of this literature that underlies the class of models developed in Section 10.8. The econometric methodology employed for nearly all of the applications in our treatment embodies much of the standard results for regression style modeling *beyond linear regression*. Some important general econometric results are examined in Section 10.7. Finally, Section 10.8 presents a class of discrete choice models, the *Hybrid Choice Model*, which combines many features of qualitative variable modeling. Some conclusions are drawn in Section 10.9.

10.2 Models of Covariation and Discrete Outcomes

Regression and regression-like models. Quantitative outcomes.[1]
Conditional means and variances. Measured Effects.
General modeling framework for qualitative variables. Di Maria, Ferreira and Lazarova (2010)

10.3 Qualitative Variables

Nonquantitative variables. Discrete and unobserved variables.
10.3.1 Discrete independent variables. Dummy variables. Factor variables.
10.3.2 Discrete dependent variables. Methodology. Indicators of outcomes.
Modeling framework. Contrast with quantitative dependent variables. Ordering.
10.3.3 Unobserved variables. Continuous, standard normal, interpretation.
10.3.4 An unobserved latent variable regression model. Constructed regressor. (M and T)

10.4 Binary Response Models

The bioassay origin story of the probit model. Antecedent of random utility.
Threshold model for discrete response. Bliss's method of probits for dosage response.

10.5 Discrete Response Variables

Generalities for discrete response. Likelihood estimation
10.5.1 Binary response. Probit and logit models. Technical details. MLE.
10.5.2 Zombie Econometrics: The linear probability model.
10.5.3 Wither heteroskedasticity? Robust estimation and inference. Cluster correction.
10.5.4 Random utility models: Probit and logit. Functional forms. Index function models.
10.5.5 A missing variable model – the EM algorithm.
10.5.6 Endogeneity: Continuous regressor. Correlations of unobservables
10.5.7 Endogeneity: Binary regressor. Recursive bivariate probit model
10.5.8 Multinomial ordered choice models. Health, well-being, happiness
10.5.8.1 Factor variables (education levels, partial effects)
10.5.8.2 Loss of generality – the generalized model and parallel regressions
10.5.9 Multinomial unordered choice modeling.

[1] This will be the first footnote

10.5.9.1 Multinomial logit. Random utility.
10.5.9.2 Willingness to pay.

10.6 Models for Continuous Latent Variables

Latent variables in response models.
10.6.1 Factor analysis, principal components, data reduction forms.
10.6.2 MIMIC model. Identification.
10.6.3 Latent variable models for multivariate normality, SEM. Multivariate models.

10.7 Econometric Methodology

equation generator. Using MathType in Word then translating equation objects to LaTeX and transporting them to Overleaf document.

10.7.1 Nonlinearity, methodology, nonlinear partial effects, optimization, Delta method.
10.7.1.1 Probit and Logit models. Others, functional form issue. Credible model.
10.7.1.2 Multinomial logit and functional form.
10.7.2 Identification by nonlinear functional form. The RBP model.
10.7.3 Aspects of Maximum Likelihood estimation
10.7.3.1 LIML/FIML: bi- and multivariate probit
10.7.3.2 Multinomial probit. Normalization.
10.7.3.3 FIML vs. LIML in new directions. Binarization in ordered choice.
10.7.4 Quasi-MLE; bivariate ordered probit.
10.7.5 Minimum distance estimation. Merging attitude information; a latent variable model.
10.7.6 GMM Estimation
10.7.7 Maximum simulated likelihood – multinomial probit
10.7.8 Panel data
10.7.8.1 REM, quadrature and simulation. Multinomial probit
10.7.8.2 FEM, the incidental parameters problem
10.7.8.3 FEM and conditional estimation, binary logit
10.7.8.4 Dynamic models and state persistence vs. time series. probit.
10.7.8.5 Jones et al OP, Greene and Rhine banking
10.7.9 Random parameters models. Latent interaction

10.8 Hybrid Choice Models

10.8.1 Integrated discrete choices / latent variables model

10.8.2 One and Two step estimation. Selectivity model

10.8.2.1 Efficiency

10.8.3 Constrained and FIML Estimation

10.9 Summary and Conclusions

References

Bliss, C. (1934). The method of probits. *Science*, *79*, 38-39.

Di Maria, C., Ferreira, S. & Lazarova, E. (2010). Shedding light on the light bulb puzzle: The role of attitudes and perceptions in the adoption of energy efficient light bulbs. *Scottish Journal of Political Eonomy*, *57*(1), 48-68.

Greene, W. H. & Hensher, D. A. (2010). *Modeling ordered choices*. Cambridge University Press.

Greene, W. H. & Hensher, D. A. (2026). Hybrid choice models in transport research. *Foundations and Trends in Econometrics*, *35*(2), 1-175.

Contoyannis, Rice, Jones et al., ca 2010, Dynamics of Health, Ordered probit
Greene and Rhine, ca 2000, Dynamic bivariate, recursive probit
Contoyannis, Jones, Rice ca 2004, Recursive bivariate probit
Williams, Generalized ordered probit
Angrist and Pischke, 2005, zombie Econometrics, Linear probability model
Muris binarization, Ordered probit
partial effects for factors
Butler and Moffitt, 2000 Quadrature and Simulation, RE probit
World Happiness report
Fama – French CAPM
More references to come...