Statistica Sinica **35** (2025), 1-25
doi:https://doi.org/10.5705/ss.02023.0231

GOODNESS OF FIT CHECKING FOR STUKEL GENERALIZED LOGISTIC REGRESSION MODELS

Yu Xia[1], Xinmin Li[2], Feifei Chen[3] and Hua Liang*[1]

[1]*George Washington University,* [2]*Qingdao University and* [3]*Beijing Normal University*

Abstract: As a useful generalization, Stukel generalized logistic models can fit binary response variable more flexibly and conveniently comparing to other generalizations. In this paper, we propose a projection-based test to check Stukel models. The test is shown to be consistent and can detect root-n local alternative hypotheses, and can be used to check the standard logistic models as well. We establish the asymptotic distribution of the proposed test under the null hypothesis and analyze asymptotic properties under the local and global alternatives. We evaluate the finite-sample performance via simulation studies and apply the proposed method to analyze a real dataset as an illustration.

Key words and phrases: Consistent test, generalized logistic models, marked empirical process, projection.

1. Introduction

The logistic regression model is often used when a binary response is to be regressed upon one or more explanatory variables. For example, this may occur when the response represents the survival or death of a patient, and the explanatory variable might be the characteristics of the individual or various treatment methods. The relationship between the response probability and dosage is often modeled using a logistic distribution function. Estimation and inference based on the maximum likelihood estimator (MLE) in logistic regression have been extensively studied in theory and widely used in social science, finance industry, and medical science (Hosmer and Lemeshow, 1989; Lindsey, 1997). See comprehensive introductions and techniques for fitting the logistic models in McCullagh and Nelder (1989) and Nelder and Wedderburn (1972).

The standard logistic model assumes the response y with mean μ, which depends on the p explanatory variables $X = (x_1, \ldots, x_p)^\top$ in the form $\mu(\eta) = \exp(\eta)/\{1+\exp(\eta)\}$, where $\eta = X^\top\boldsymbol{\beta}$ and $\boldsymbol{\beta}$ is a p-vector of unknown parameters in $\mathbf{R}^p$. The model still has its limitations, though its popularity is irreproachable. As Stukel (1988) pointed out that the form of $\mu(\eta)$ restricts a symmetric pattern about $\eta = 0$. When data are not symmetric, or even symmetric but have a steeper or gentler incline in the central probability region, the logistic model may

*Corresponding author. E-mail: hliang@gwu.edu