Available online at www.sciencedirect.com
ScienceDirect
Economics Letters 98 (2008) 176–184
economics letters
www.elsevier.com/locate/econbase

Testing for discrete choice models

Xu Zheng *

School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China

Received 25 November 2005; received in revised form 21 April 2007; accepted 25 April 2007
Available online 3 May 2007

Abstract

This paper proposes a new specification test for discrete choice models based on nonparametric estimation of the likelihood ratio. The test is robust against any possible misspecification. A bootstrap procedure is proposed to obtain critical values in small samples. Monte Carlo simulations show that the proposed test can have much higher power than some of the well known parametric tests.

Keywords: Discrete choice models; Specification test; Nonparametric method; Likelihood ratio test; Multinomial logit

JEL Classification: C1; C4

1. Introduction

Discrete choice models are widely used in economics and other social sciences. The most popular ones are the binary logit and probit models and the multinomial logit model of McFadden (1973). Tests for the multinomial logit model have been proposed by Hausman and McFadden (1984), Hsiao and Small (1985), McFadden (1987), and Wills (1987). Inferences for general discrete choice models can be conducted using the classical trinity: Wald, likelihood ratio (LR), and Lagrange multiplier tests. It is often easiest to construct a LR test once an alternative is specified. As shown in Wills (1987), the test of Hausman and McFadden (1984) has the same asymptotic local power as the classical tests. However, the performances of these parametric tests rely on correct specification of the alternative model. If the alternative is misspecified, those tests may have low or no power.

* Tel.: +86 21 65904613; fax: +86 21 65904198.
E-mail address: xuzheng1@gmail.com.

0165-1765/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2007.04.027

This paper proposes a new test for general discrete choice models. Unlike the aforementioned tests, the proposed test does not assume a parametric alternative and thus does not suffer the non-robustness problem associated with parametric tests. The test is in the spirit of LR test and based on comparing the parametric null model with the nonparametrically estimated alternative model. The test is shown to be more powerful than the parametric tests and also easy to calculate. Testing a parametric model against a nonparametric alternative has been extensively studied in the literature. The related work includes Aït-Sahalia et al. (2001), Bickel et al. (2006), Bierens (1990), Chen and Fan (1999), Chen et al. (2005), Ellison and Fisher-Ellison (2000), Fan and Li (1996), Hardle and Mammen (1993), Hong and White (1995), Stinchcombe and White (1998), Wooldridge (1992), Yatchew (1992), Zheng (1996, 1998a,b, 2000) and others. Most of these tests are designed for testing parametric regression models and some are testing more targeted alternatives, such as Chen et al. (2005) and Bickel et al. (2006). Some of these tests, such as Zheng's (2000), assume a continuous dependent variable and thus do not apply to the multinomial model. Like most of these tests, the test of this paper also considers a general nonparametric alternative and is designed to have power against any direction of departure from the null model. The test may be interpreted as a nonparametric LR test. The plan of the paper is as follows. Section 2 derives the nonparametric test and provides the asymptotic behavior of the test. Section 3 provides a bootstrap procedure and some Monte Carlo simulations.

2. The derivation of the test

Consider a multinomial model where the dependent variable Y takes on $m+1$ discrete values 0, 1, … , m. Denote the conditional p.f. of Y given the covariates X by $p(y|x)$. X takes on values in R^d. Let $p_1(x)$ be the marginal d.f. of X. Let $\{(y_i, x_i)\}_{i=1}^n$ be a random sample of (Y, X). In parametric models, $p(y|x)$ is assumed to belong to a parametric family of known real-valued functions $f(y|x,\theta)$ on $R \times R^d \times \Theta$, where $\Theta \subset R^k$. The null hypothesis to be tested is that the parametric model is correct: H_0: $p(y|x)=f(y|x,\theta_0)$ for some $\theta_0 \in \Theta$, while the alternative is that the null is false: H_1: $p(y|x) \neq f(y|x,\theta)$ for any $\theta \in \Theta$. Note that the alternative consists of all the possible alternative models. A test is consistent if its asymptotic power is 1. Let $\hat{\theta}$ be the MLE of θ_0 under the null. Since the alternative is not specified, $p(y|x)$ is estimated by a nonparametric estimator $\hat{p}(y|x)$. Then a nonparametric LR type of test may be based on

$$\mathrm{NLR}_1 \equiv 2\left[\sum_{i=1}^{n} \ln(\hat{p}(y_i|x_i)) - \sum_{i=1}^{n} \ln(f(y_i|x_i,\hat{\theta}))\right] = 2\sum_{i=1}^{n} \ln\left[\frac{\hat{p}(y_i|x_i)}{f(y_i|x_i,\hat{\theta})}\right]. \tag{1}$$

However, it is difficult to establish the asymptotic distribution of NLR_1 since it is a nonlinear function of $\hat{p}(y_i|x_i)$. Since $\ln(x) \approx x-1$, we take a first-order linear expansion of NLR_1 and weigh it by an estimator $\hat{p}_1(x)$ of $p(x)$ to get

$$\mathrm{NLR} \equiv 2\sum_{i=1}^{n} \hat{p}_1(x_i)\left[\frac{\hat{p}(y_i|x_i) - f(y_i|x_i,\hat{\theta})}{f(y_i|x_i,\hat{\theta})}\right]. \tag{2}$$

Denote $1(A)=1$ if event A occurs and 0 otherwise. $p(y|x)$ and $p(x)$ can be estimated by a kernel type of estimator

$$\hat{p}(y|x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{h^d} 1(y=y_i) K\left(\frac{x-x_j}{h}\right) / \hat{p}_i(x), \tag{3}$$