Available online at www.sciencedirect.com

ScienceDirect

Journal of Econometrics 143 (2008) 37–55

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Distribution-free specification tests of conditional models

Miguel A. Delgado[a,*], Winfried Stute[b]

[a]Departamento de Economía, Universidad Carlos III de Madrid, C./ Madrid 126-128, Getafe 28903, Madrid, Spain
[b]Mathematical Institute, University of Giessen, Arndtstr. 2, Germany

Available online 19 August 2007

Abstract

This article proposes a class of asymptotically distribution-free specification tests for parametric conditional distributions. These tests are based on a martingale transform of a proper sequential empirical process of conditionally transformed data. Standard continuous functionals of this martingale provide omnibus tests while linear combinations of the orthogonal components in its spectral representation form a basis for directional tests. Finally, Neyman-type smooth tests, a compromise between directional and omnibus tests, are discussed. As a special example we study in detail the construction of directional tests for the null hypothesis of conditional normality versus heteroskedastic contiguous alternatives. A small Monte Carlo study shows that our tests attain the nominal level already for small sample sizes.

Keywords: Conditional models; Martingale transformation; Sequential empirical process; Specification tests

1. Introduction

The correct specification of a statistical model is important for several reasons. First, it provides a convenient framework to describe and understand, for example, the dynamics of a time series or a causal relation between independent and dependent variables in regression. In each case it turns out that conditional quantities like autoregressive functions or conditional distributions are of major interest, while marginal distributions of explanatory variables may be considered as parametric or nonparametric nuisance parameter functions. The choice of the model has some consequences on the estimation of unknown parameters and hence on the interpretation of data or the prediction of future values of a dependent variable. The validity of statistical inferences based on conditional maximum likelihood principle, e.g., relies on the correct specification of the conditional distribution model. In particular, the popular Lagrange multiplier and likelihood ratio tests on parameter restrictions are invalid under misspecification, though robust but inefficient inferences are possible. However, classical procedures are optimal under a correct specification. Applications using conditional maximum likelihood are available in abundant supply in economics, as well as in any other disciplines where statistical inference is indispensable. The correct specification of conditional distributions is especially crucial in microeconometrics and biostatistics, where parameter identification is sustained by a correct specification. In these cases, parameter estimates are inconsistent under misspecification. See the

*Corresponding author.
E-mail addresses: miguelangel.delgado@uc3m.es (M.A. Delgado), winfried.stute@math.uni-giessen.de (W. Stute).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.08.006

classical monograph by Maddala (1983) on limited-dependent and qualitative variables models, Cameron and Trivedi (1998) for count data models, or Lancaster (1990) for duration models.

In the simple case of independent identically distributed observations the history of goodness-of-fit tests started with the classical χ^2-test for cell probabilities. For continuous variables most of the procedures, like Kolmogorov–Smirnov and Cramér–von Mises tests, are based on proper functionals of the empirical process. When the model to be tested is composite, the need to estimate unknown parameters has some impact on the distributional character under the null model so that available tables of critical values are no longer valid. See the work of Gikhman (1953) and Kac et al. (1955) for some early fundamental contributions in this context. A formal derivation of the limit process is due to Durbin (1973) and Neuhaus (1973, 1976), among others. For practical purposes, critical values of the tests can be obtained either through resampling or through the orthogonal components in the spectral representation of the underlying empirical process, as suggested by Durbin et al. (1975).

A different approach was initiated by Khmaladze (1981), who proposed to transform the empirical process to an appropriate martingale, which in distribution may then be approximated by a time-transformed Brownian Motion. As a consequence, classical functionals of these processes like the Kolmogorov–Smirnov or Cramér–von Mises test statistics become asymptotically distribution-free so that existing tables can be used.

In this paper we are interested, for a multivariate observation (X, Y), in the conditional distribution of Y given $X = x$. For the related question of testing just the conditional mean and not the whole conditional distributional structure, the literature is much more elaborate. Härdle and Mammen (1993) were among the first to compare parametric and nonparametric fits. These tests require some smoothing to the effect that the power of these tests may depend on the choice of the smoothing parameter. Stute (1997) investigated so-called integrated regression function (or cusum) processes which avoid smoothing and at the same time allow for a principal component analysis. If we replace (in our notation) Y by indicators $1_{\{Y \leqslant y\}}$, these approaches lead to tests of conditional probability models and may be found in Andrews (1997). In particular he investigated the Kolmogorov–Smirnov test. Due to the complicated distributional character of the test statistic, a bootstrap approximation was proposed and studied. The martingale transformation of the cusum process for fixed design and linear regression is due to Brown et al. (1975). The random design case with a possibly nonlinear regression function has been dealt with in Stute et al. (1998), while applications to time series and generalized linear models may be found in Koul and Stute (1999) and Stute and Zhu (2002). See also Nikabadze and Stute (1997) and Khmaladze and Koul (2004). Zheng (2000) has extended the smoothing approach to specification tests of conditional distributions, while Bai (2003) has applied Khmaladze's martingale approach to tests of the marginal distribution of time series innovations.

To motivate the approach of the present paper we recall a fundamental result due to Rosenblatt (1952). Namely, let (X, Y) be a bivariate random vector with an unknown continuous distribution function F. Denote with F_X the marginal distribution function of X and let $F_{Y|X}(y|x)$ be the conditional distribution function of Y given $X = x$ evaluated at y. Given F_X, F is uniquely determined through $F_{Y|X}$ and vice versa.

In nonparametric testing for F, it is known that tests based on the empirical distribution function are no longer distribution-free. In this context, Rosenblatt (1952) used F_X and $F_{Y|X}$ to introduce a transformation $T = T(X, Y) = (U, V)$ of (X, Y), which maps (X, Y) into a vector (U, V) such that U and V are independent and uniformly distributed on [0,1]. Just put $U = F_X(X)$ and $V = F_{Y|X}(Y|X)$. It is easy to recover (X, Y) from (U, V). Actually, we have with probability one $(X, Y) = (F_X^{-1}(U), F_{Y|X}^{-1}(V|F_X^{-1}(U))$, where G^{-1} denotes the quantile function of a distribution function G. The transformation T can be extended to higher dimensions, but in this paper, for most of the time, we shall stick to the bivariate case. We rather study the important situation when $X = Z^{\mathrm{T}}\delta_0$, for a $p \times 1$ random vector Z and an unknown parameter vector δ_0, so that the multidimensionality of the model enters through a proper projection of a random vector Z. The extension to the case where $X = m(Z, \delta_0)$ for a suitably smooth m is routine. These so-called dimension reducing models are popular in applied fields and naturally lead to an input–output analysis in which, at an intermediate step, the independent variable is univariate. This is relevant in many econometric applications, where one assumes a regression model with innovations independent of the explanatory variables, e.g., limited-dependent variable models.

The Rosenblatt transform T constitutes the extension of the transformation $U = F_X(X)$, which is basic in the analysis of univariate data and leads to many distribution-free procedures based on ranks or