



HAL
open science

On specification testing ordered discrete choice models

Juan Mora, Ana I. Moro-Egido

► **To cite this version:**

Juan Mora, Ana I. Moro-Egido. On specification testing ordered discrete choice models. *Econometrics*, 2008, 143 (1), pp.191. 10.1016/j.jeconom.2007.08.004 . hal-00501801

HAL Id: hal-00501801

<https://hal.science/hal-00501801>

Submitted on 12 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

On specification testing ordered discrete choice models

Juan Mora, Ana I. Moro-Egido

PII: S0304-4076(07)00163-7
DOI: doi:10.1016/j.jeconom.2007.08.004
Reference: ECONOM 2973

To appear in: *Journal of Econometrics*

Cite this article as: Juan Mora and Ana I. Moro-Egido, On specification testing ordered discrete choice models, *Journal of Econometrics* (2007), doi:10.1016/j.jeconom.2007.08.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jeconom

On Specification Testing of Ordered Discrete Choice Models

by JUAN MORA* and ANA I. MORO-EGIDO†

July, 2007‡

Abstract

We discuss how to test the specification of an ordered discrete choice model against a general alternative. Two main approaches can be followed: tests based on moment conditions and tests based on comparisons between parametric and nonparametric estimations. Following these approaches, various statistics are proposed and their asymptotic properties are discussed. The performance of the statistics is compared by means of simulations. An easy-to-compute variant of the standard moment-based statistic yields the best results in models with a single explanatory variable. In models with various explanatory variables the results are less conclusive, since the relative performance of the statistics depends on both the fit of the model and the type of misspecification that is considered.

J.E.L. Classification Numbers: C25, C52, C15.

Keywords: Specification Tests; Ordered Discrete Choice Models; Simulation.

*Corresponding Author; Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Apartado de Correos 99, 03080 Alicante, Spain. E-mail address: juan@merlin.fae.ua.es.

†Departamento de Teoría e Historia Económica, Universidad de Granada, Campus Cartuja s/n, 18011 Granada, Spain. E-mail address: aimoro@ugr.es.

‡We are grateful to F. Peñaranda and two anonymous referees for their comments.

Ordered discrete choice variables often appear in Statistics and Econometrics as a dependent variable. Examples of this type of dependent variables used in recent applied works include: educational level attained by individuals, female labour participation (work full-time/work part-time/not to work), and level of demand for a new product or service. The outcomes of an ordered discrete choice variable Y are usually labelled as $0, 1, \dots, J$. Given certain explanatory variables $X = (X_1, \dots, X_k)'$, the researcher is usually interested in analysing whether the proposed explanatory variables are significant or not, and/or providing accurate estimates of the conditional probabilities $\Pr(Y = j \mid X = x)$. The most frequently used parametric model for an ordered discrete choice variable arises when one assumes the existence of a latent continuous dependent variable Y^* for which a linear regression model $Y^* = X'\beta_0 + u$ holds; the non-observed variable Y^* and the observed variable Y are assumed to be related as follows: $Y = j$ if $\mu_{0,j-1} \leq Y^* < \mu_{0j}$, for $j = 0, 1, \dots, J$, where $\mu_{0,-1} \equiv -\infty$, $\mu_{0,J} \equiv +\infty$ and $\mu_{00}, \mu_{01}, \dots, \mu_{0,J-1}$ are threshold parameters such that $\mu_{00} \leq \mu_{01} \leq \dots \leq \mu_{0,J-1}$. Assuming independence between u and X , this implies that

$$\Pr(Y = j \mid X) = F(\mu_{0j} - X'\beta_0) - F(\mu_{0,j-1} - X'\beta_0), \quad \text{for } j = 0, 1, \dots, J, \quad (1)$$

where $F(\cdot)$ is the cumulative distribution function (cdf) of u , usually referred to as the “link function”. To ensure identification μ_{00} is usually assumed to be 0; additionally, it is assumed that $F(\cdot)$ is entirely known, and typical choices are the standard normal cdf (“ordered probit”) and the logistic cdf (“ordered logit”). With these assumptions a full parameterization of the conditional distribution $Y \mid X = x$ is obtained, with parameter vector $(\beta_0', \mu_0')' \subset \mathbb{R}^{k+J-1}$. Parameter estimates and predicted probabilities are inconsistent if these assumptions are not met. Therefore, it is especially important to test the null hypothesis that the parametric specification (1) is correct. The aim of

this paper is to describe and compare procedures to perform this test.

The usual approach to test one (or some) of the assumptions of a parametric ordered discrete choice model is to construct test statistics based on one (or various) moment conditions derived from (1), using the methodology described in Newey (1985) and Tauchen (1985). For example, Skeels and Vella (1999) derive various statistics for the probit model ($J = 1$); and Butler and Chatterjee (1997) propose a test of overidentifying restrictions that can be used if $J \geq 2$. However, most of the recent research has focused on the derivation of moment conditions that allow to detect specific departures from the proposed specification; thus, Weiss (1997) proposes to test if u is homoskedastic against some heteroskedastic alternatives; Johnson (1996) proposes to test if u is normal against the alternative that it is a member of the Pearson family; and Murphy (1996) proposes to test if $F(\cdot)$ is logistic against various alternatives. In contrast, we focus on omnibus specification tests that do not address to any specific alternative.

The starting point of the moment-based statistics that we consider here is the set of simplest moment conditions that are derived from (1), namely, the expectations of $\mathbf{1}_{\{Y=j\}}$, for $j = 0, 1, \dots, J$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. We discuss how asymptotically chi-square statistics can be derived from these moment conditions. We consider various moment-based statistics that only differ in how the covariance matrix of the sample moments is estimated, but it is well-known that this estimation plays a crucial role in the small-sample performance of the statistic. This phenomenon was thoroughly studied by Orme (1990), who suggests to use covariance matrix estimators based on conditional expectations of analytical derivatives to improve small-sample performance. Here we can follow this suggestion since the specification of our model allows us to obtain any conditional expectation (note that this makes our approach different to that in Skeels and Vella 1999, who estimate the covariance matrix using analytical derivatives without taking expectations). Additionally we prove that, as

many others moment-based statistics (see e.g. Wooldridge 1990), the statistic obtained with a conditional-expectation-based covariance matrix can be computed using an artificial regression. Finally, following Andrews (1988), we describe how the number of moments on which the statistics are based can be increased by partitioning the support of the regressors, and discuss whether this leads to an increase in power.

When a finite set of unconditional moment restrictions is used to test a null hypothesis that specifies a conditional moment restriction, the resulting test statistic is in general not consistent against any possible alternative, since it might happen that all the unconditional moment restrictions that are being tested hold, but the null hypothesis does not. For this reason, the statistics described above are not consistent. Andrews (1997) proposes a consistent statistic to test a parametric specification of a conditional cdf that can be used in our context; but most of the recent literature on consistent specification tests has focused on regression models. When $J > 1$, the specification that we consider is not a regression model but the methodology that is applied in regression models can also be applied in our context: this will be our strategy to derive other consistent specification tests for ordered discrete choice models. Most of the consistent specification tests for regression models that have appeared in the literature are derived by comparing parametric and nonparametric regression estimates. Many of them require the use of a smoothing value; this is an undesirable property, since the choice of the smoothing value plays a key role in the results, and the problem of how to choose it is far from trivial. For this reason, here we only consider statistics whose behavior does not depend crucially on the choice of a smoothing value. More specifically, we consider the statistics proposed in Horowitz and Spokoiny (2001), Stute and Zhu (2002) and Guerre and Lavergne (2005), and adapt them to our context.

We use a variety of Monte Carlo experiments to analyse the performance of the statistics and to compare the relative merits of all of them. The experiments are

designed to cover the most typical departures from the null hypothesis that one would like to detect in practice: non-linearity, heteroskedasticity and misspecification of the cdf of the error in the latent regression model.

The rest of the paper is organized as follows. In Sections 2 and 3 the test statistics are derived. In Section 4 we describe the Monte Carlo experiments, discuss their results and present an empirical application. In Section 5 we conclude. Some technical details are relegated to an Appendix.

2. STATISTICS BASED ON MOMENT CONDITIONS

We assume that independent and identically distributed (i.i.d.) observations $(Y_i, X_i)'$ are available, where, hereafter, $i = 1, \dots, n$. The following notation will be used: $D_{ji} \equiv \mathbf{1}_{\{Y_i=j\}}$, for $j = 0, 1, \dots, J$; and, given $\theta \equiv (\beta', \mu')' \in \Theta \subset \mathbb{R}^{k+J-1}$, $p_{0i}(\theta) \equiv F(-X_i'\beta)$; $p_{Ji}(\theta) \equiv 1 - F(\mu_{J-1} - X_i'\beta)$; if $J \geq 2$, $p_{1i}(\theta) \equiv F(\mu_1 - X_i'\beta) - F(-X_i'\beta)$; and if $J \geq 3$, $p_{ji}(\theta) \equiv F(\mu_j - X_i'\beta) - F(\mu_{j-1} - X_i'\beta)$, for $j = 2, \dots, J-1$.

Define $m_{ji}(\theta) \equiv D_{ji} - p_{ji}(\theta)$. It follows from (1) that $E\{m_{ji}(\theta_0)\} = 0$, for $j = 0, 1, \dots, J$. This yields $J+1$ moment conditions but, as the sum of all probabilities adds to one, only J are used to construct a test statistic. Specifically, we consider the random vector $\sum_{i=1}^n \mathbf{m}_i(\hat{\theta})$, where $\mathbf{m}_i(\theta)$ is the $J \times 1$ column vector whose j -th component is $m_{ji}(\theta)$ and $\hat{\theta}$ is a well-behaved estimator of θ_0 . To derive an asymptotically valid test statistic, note that using a first-order Taylor expansion it follows that $n^{-1/2} \sum_{i=1}^n \mathbf{m}_i(\hat{\theta}) = n^{-1/2} \sum_{i=1}^n \mathbf{m}_i(\theta_0) + \mathbf{B}_0 \times n^{1/2}(\hat{\theta} - \theta_0) + o_p(1)$, where $\mathbf{B}_0 \equiv E\{\mathbf{B}_i(\theta_0)\}$ and $\mathbf{B}_i(\theta)$ denotes the $J \times (k+J-1)$ matrix whose j -th row is $\partial m_{ji}(\theta) / \partial \theta'$. In our context, the natural way to estimate θ_0 is maximum likelihood (ML). Assuming that certain regularity conditions hold, the ML estimator $\hat{\theta}$ satisfies that $n^{1/2}(\hat{\theta} - \theta_0) = \mathbf{A}_0^{-1} \times n^{-1/2} \sum_{i=1}^n \mathbf{g}_i(\theta_0) + o_p(1)$, where $\mathbf{g}_i(\theta) \equiv \sum_{j=1}^J D_{ji} \partial \ln p_{ji}(\theta) / \partial \theta$ is the derivative with respect to θ of the i -th term in $\ln L(\theta)$, and $\mathbf{A}_0 = E\{\mathbf{A}_i(\theta_0)\}$,

for $\mathbf{A}_i(\theta) \equiv -\partial \mathbf{g}_i(\theta) / \partial \theta'$, is the limiting information matrix. Inserting the asymptotic expansion of $n^{1/2}(\hat{\theta} - \theta_0)$ into the Taylor expansion of $n^{-1/2} \sum_{i=1}^n \mathbf{m}_i(\hat{\theta})$ it follows that

$$n^{-1/2} \sum_{i=1}^n \mathbf{m}_i(\hat{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_0), \quad (2)$$

where $\mathbf{V}_0 \equiv [\mathbf{I}_J : \mathbf{B}_0 \mathbf{A}_0^{-1}] \mathbf{Q}_0 [\mathbf{I}_J : \mathbf{B}_0 \mathbf{A}_0^{-1}]'$, \mathbf{I}_J is the $J \times J$ identity matrix, $\mathbf{Q}_0 \equiv E\{\mathbf{Q}_i(\theta_0)\}$ and $\mathbf{Q}_i(\theta) \equiv (\mathbf{m}_i(\theta)', \mathbf{g}_i(\theta)')'(\mathbf{m}_i(\theta)', \mathbf{g}_i(\theta)')$. To derive a test statistic, a consistent estimator of \mathbf{V}_0 must be proposed. It is worthwhile discussing in detail how this can be done, since it is well-known (Orme 1990) that the finite-sample performance of moment-based statistics crucially depends on this.

The natural candidate for estimating \mathbf{V}_0 is $\mathbf{V}_{n,1} \equiv [\mathbf{I}_J : \mathbf{B}_n \mathbf{A}_n^{-1}] \mathbf{Q}_n [\mathbf{I}_J : \mathbf{B}_n \mathbf{A}_n^{-1}]'$, where $\mathbf{Q}_n \equiv n^{-1} \sum_{i=1}^n \mathbf{Q}_i(\hat{\theta})$, $\mathbf{B}_n \equiv n^{-1} \sum_{i=1}^n \mathbf{B}_i(\hat{\theta})$, $\mathbf{A}_n \equiv n^{-1} \sum_{i=1}^n \mathbf{A}_i(\hat{\theta})$. But it is possible to derive an alternative estimator of \mathbf{V}_0 that leads to a computationally simpler statistic: the information matrix equality ensures that $E\{\mathbf{g}_i(\theta_0) \mathbf{g}_i(\theta_0)'\} = \mathbf{A}_0$, and it is easy to check that $E\{\mathbf{m}_i(\theta_0) \mathbf{g}_i(\theta_0)'\} = -\mathbf{B}_0$; hence it follows that \mathbf{V}_0 equals

$$E\{\mathbf{m}_i(\theta_0) \mathbf{m}_i(\theta_0)'\} - E\{\mathbf{m}_i(\theta_0) \mathbf{g}_i(\theta_0)'\} E\{\mathbf{g}_i(\theta_0) \mathbf{g}_i(\theta_0)'\}^{-1} E\{\mathbf{g}_i(\theta_0) \mathbf{m}_i(\theta_0)'\}. \quad (3)$$

Thus we consider $\mathbf{V}_{n,2} \equiv n^{-1} [\sum_{i=1}^n \mathbf{m}_i(\hat{\theta}) \mathbf{m}_i(\hat{\theta})' - \sum_{i=1}^n \mathbf{m}_i(\hat{\theta}) \mathbf{g}_i(\hat{\theta})' \{ \sum_{i=1}^n \mathbf{g}_i(\hat{\theta}) \mathbf{g}_i(\hat{\theta})' \}^{-1} \sum_{i=1}^n \mathbf{g}_i(\hat{\theta}) \mathbf{m}_i(\hat{\theta})']$. Note that $\mathbf{V}_{n,1}$ and $\mathbf{V}_{n,2}$, obtained by simply replacing population moments by sample moments, are the standard estimates of \mathbf{V}_0 following the Newey-Tauchen methodology. However, in our context we can do better than that: our null hypothesis specifies the conditional distribution $Y \mid X = x$; hence, any conditional expectation can be derived, and the sample analog of the conditional expectation is, by the law of iterated expectations, a consistent estimator of the population moment. This approach was proposed by Orme (1990) in the context of information matrix tests. Here, (3) suggests that we can estimate \mathbf{V}_0 with $\mathbf{V}_{n,3} \equiv n^{-1} \sum_{i=1}^n \mathbf{V}_{i,3}(\hat{\theta})$, where $\mathbf{V}_{i,3}(\theta) \equiv E_X\{\mathbf{m}_i(\theta) \mathbf{m}_i(\theta)'\} - E_X\{\mathbf{m}_i(\theta) \mathbf{g}_i(\theta)'\} E_X\{\mathbf{g}_i(\theta) \mathbf{g}_i(\theta)'\}^{-1} E_X\{\mathbf{g}_i(\theta) \mathbf{m}_i(\theta)'\}$.

To sum up, we can consider three possible estimates for \mathbf{V}_0 and thus derive three pos-

sible test statistics $C_{n,l}^{(M)} \equiv n^{-1} \{ \sum_{i=1}^n \mathbf{m}_i(\theta)' \} \mathbf{V}_{n,l}^- \{ \sum_{i=1}^n \mathbf{m}_i(\theta) \}$, for $l = 1, 2, 3$, where $\mathbf{V}_{n,l}^-$ denotes the Moore-Penrose inverse of $\mathbf{V}_{n,l}$, which is, under certain conditions, a consistent estimator of the Moore-Penrose inverse of \mathbf{V}_0 (see Andrews 1987). From (2) it follows that if (1) holds then $C_{n,l}^{(M)} \xrightarrow{d} \chi_{\text{rk}(\mathbf{V}_0)}^2$, what justifies the use of $C_{n,l}^{(M)}$ as an asymptotically valid test statistic. Typically, $\text{rk}(\mathbf{V}_0) = J$ and $\mathbf{V}_{n,l}^-$ is just the inverse of $\mathbf{V}_{n,l}$; but if one suspects that \mathbf{V}_0 might not be invertible (this happens e.g. if X_1 is constant and $\beta_{02} = \dots = \beta_{0k} = 0$), it would be desirable to test whether the hypothesis $\text{rk}(\mathbf{V}_0) = J$ is plausible (see e.g. Robin and Smith 2000).

The computation of the three statistics can be made using the analytical expressions that are given in the Appendix. But $C_{n,2}^{(M)}$ and $C_{n,3}^{(M)}$ can also be computed using artificial regressions. On the one hand, $C_{n,2}^{(M)}$ can be computed as the explained sum of squares (ESS) in the artificial regression of a vector of ones on $\mathbf{m}_i(\hat{\theta})'$ and $\mathbf{g}_i(\hat{\theta})$ (see Newey 1985 and Tauchen 1985). On the other hand, if we denote $\hat{p}_{ji} \equiv p_{ji}(\hat{\theta})$ and $\hat{\delta}_{ji} \equiv 1 - F(\hat{\mu}_j - X_i' \hat{\beta}) + F(-X_i' \hat{\beta})$, and consider the $J \times 1$ vectors $\hat{\mathbf{c}}_{ji}$, $\hat{\mathbf{d}}_{ji}$, $\hat{\mathbf{e}}_{i,}$, $\hat{\mathbf{f}}_{ji}$, whose l -th components are defined by $\hat{\mathbf{c}}_{ji,l} \equiv \{ \mathbf{1}_{\{l < j\}} \hat{p}_{ji} + \mathbf{1}_{\{l=j\}} \hat{\delta}_{j-1,i} \} / (\hat{p}_{ji} \hat{\delta}_{ji} \hat{\delta}_{j-1,i})^{1/2}$, $\hat{\mathbf{d}}_{ji,l} \equiv \hat{p}_{li}^{1/2} [-\mathbf{1}_{\{l < j\}} \hat{p}_{ji} + \mathbf{1}_{\{l=j\}} \hat{\delta}_{ji}] / (\hat{\delta}_{li} \hat{\delta}_{l-1,i})^{1/2}$, $\hat{\mathbf{e}}_{i,l} \equiv f(\hat{\mu}_l - X_i' \hat{\beta}) - f(\hat{\mu}_{l-1} - X_i' \hat{\beta})$ and $\hat{\mathbf{f}}_{ji,l} \equiv \{ \mathbf{1}_{\{l=j+1\}} - \mathbf{1}_{\{l=j\}} \} f(\hat{\mu}_j - X_i' \hat{\beta})$, then it is possible to prove that $C_{n,3}^{(M)}$ coincides with the ESS in the artificial regression with vector of dependent observations \mathbf{Z} and matrix of observations \mathbf{W} , where $\mathbf{Z} \equiv [\mathbf{z}'_1, \dots, \mathbf{z}'_n]'$, \mathbf{z}_i is the $J \times 1$ vector whose j -th element is $\hat{\mathbf{c}}'_{ji} \hat{\mathbf{m}}_i$, $\mathbf{W} \equiv [\mathbf{W}^{(1)} : \mathbf{W}^{(2)} : \mathbf{W}^{(3)}]$, $\mathbf{W}^{(l)} \equiv [\mathbf{w}_1^{(l)'} , \dots, \mathbf{w}_n^{(l)'}]'$ for $l = 1, 2, 3$, $\mathbf{w}_i^{(1)}$ is the $J \times J$ matrix whose j -th column is $\hat{\mathbf{d}}_{ji}$, $\mathbf{w}_i^{(2)}$ is the $J \times k$ matrix whose j -th row is $\hat{\mathbf{c}}'_{ji} \hat{\mathbf{e}}_i X_i'$ and $\mathbf{w}_i^{(3)}$ is the $J \times (J-1)$ matrix whose (j, l) element is $\hat{\mathbf{c}}'_{ji} \hat{\mathbf{f}}_{li}$. This is deduced taking into account that from the analytical expressions derived in the Appendix it follows that $\mathbf{Z}' \mathbf{W}^{(1)} = \sum_{i=1}^n \mathbf{m}_i(\hat{\theta})'$, $\mathbf{Z}' \mathbf{W}_* = - \sum_{i=1}^n \mathbf{g}_i(\hat{\theta}) = \mathbf{0}$, where $\mathbf{W}_* \equiv [\mathbf{W}^{(2)} : \mathbf{W}^{(3)}]$, and $\mathbf{W}^{(1)'} \mathbf{W}^{(1)} - \mathbf{W}^{(1)'} \mathbf{W}_* (\mathbf{W}_* \mathbf{W}_*)^{-1} \mathbf{W}_* \mathbf{W}^{(1)} = n \mathbf{V}_{n,3}$.

Andrews (1988) proposes increasing the degrees of freedom of moment-based statis-

tics by partitioning the support of the regressors. Let us assume that the support of X_i is partitioned into G subsets A_1, \dots, A_G . If we define $m_{jgi}(\theta) \equiv \mathbf{1}_{\{X_i \in A_g\}} m_{ji}(\theta)$, for $j = 1, \dots, J$ and $g = 1, \dots, G$, we can consider the JG moment conditions $E\{m_{jgi}(\theta_0)\} = 0$. To derive a test statistic, we define $\mathbf{m}_i^{(P)}(\theta) \equiv \mathbf{m}_i(\theta) \otimes \mathbf{P}_i$, where \mathbf{P}_i is the $G \times 1$ matrix whose g -th row is $\mathbf{1}_{\{X_i \in A_g\}}$, and consider $\sum_{i=1}^n \mathbf{m}_i^{(P)}(\hat{\theta})$. As above, it follows that

$$n^{-1/2} \sum_{i=1}^n \mathbf{m}_i^{(P)}(\hat{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_0^{(P)}), \quad (4)$$

where $\mathbf{V}_0^{(P)} \equiv [\mathbf{I}_{JG} : \mathbf{B}_0^{(P)} \mathbf{A}_0^{-1}] \mathbf{Q}_0^{(P)} [\mathbf{I}_{JG} : \mathbf{B}_0^{(P)} \mathbf{A}_0^{-1}]'$, $\mathbf{B}_0^{(P)} \equiv E\{\mathbf{B}_i^{(P)}(\theta_0)\}$, $\mathbf{Q}_0^{(P)} \equiv E\{\mathbf{Q}_i^{(P)}(\theta_0)\}$, $\mathbf{B}_i^{(P)}(\theta) \equiv \mathbf{B}_i(\theta) \otimes \mathbf{P}_i$ and $\mathbf{Q}_i^{(P)}(\theta) \equiv (\mathbf{m}_i^{(P)}(\theta)', \mathbf{g}_i(\theta)')' (\mathbf{m}_i^{(P)}(\theta)', \mathbf{g}_i(\theta)')$. Now, the natural estimator for $\mathbf{V}_0^{(P)}$ is $\mathbf{V}_{n,1}^{(P)} \equiv [\mathbf{I}_{JG} : \mathbf{B}_n^{(P)} \mathbf{A}_n^{-1}] \mathbf{Q}_n^{(P)} [\mathbf{I}_{JG} : \mathbf{B}_n^{(P)} \mathbf{A}_n^{-1}]'$, where $\mathbf{B}_n^{(P)} \equiv n^{-1} \sum_{i=1}^n \mathbf{B}_i^{(P)}(\hat{\theta})$ and $\mathbf{Q}_n^{(P)} \equiv n^{-1} \sum_{i=1}^n \mathbf{Q}_i^{(P)}(\hat{\theta})$. But two other estimators can be proposed: $\mathbf{V}_{n,2}^{(P)}$ and $\mathbf{V}_{n,3}^{(P)}$, defined in the same way as $\mathbf{V}_{n,2}$ and $\mathbf{V}_{n,3}$, respectively, but replacing $\mathbf{m}_i(\hat{\theta})$ by $\mathbf{m}_i^{(P)}(\hat{\theta})$. Thus we obtain three different test statistics $C_{n,l}^{(MP)} \equiv n^{-1} \{ \sum_{i=1}^n \mathbf{m}_i^{(P)}(\hat{\theta})' \} \mathbf{V}_{n,l}^{(P)} \{ \sum_{i=1}^n \mathbf{m}_i^{(P)}(\hat{\theta}) \}$, for $l = 1, 2, 3$. From (4) it follows that if (1) holds then $C_{n,l}^{(MP)} \xrightarrow{d} \chi_{\text{rk}(\mathbf{V}_0^{(P)})}^2$. Again, $C_{n,2}^{(MP)}$ and $C_{n,3}^{(MP)}$ can be computed using artificial regressions: the former coincides with the ESS in the artificial regression of a vector of ones on $\mathbf{m}_i^{(P)}(\hat{\theta})$ and $\mathbf{g}_i(\hat{\theta})$; and with a similar reasoning as above it follows that $C_{n,3}^{(MP)}$ coincides with the ESS in the artificial regression with vector of dependent observations $\mathbf{Z}^{(P)} \equiv [\mathbf{z}_1^{(P)'}, \dots, \mathbf{z}_n^{(P)'}]'$ and matrix of observations $\mathbf{W}^{(P)} \equiv [\mathbf{W}^{(P1)} : \mathbf{W}^{(P2)} : \mathbf{W}^{(P3)}]$, where $\mathbf{z}_i^{(P)} \equiv \mathbf{z}_i \otimes \mathbf{P}_i$, $\mathbf{W}^{(Pl)} = [\mathbf{w}_1^{(Pl)'}, \dots, \mathbf{w}_n^{(Pl)'}]'$ for $l = 1, 2, 3$, $\mathbf{w}_i^{(P1)} \equiv \mathbf{w}_i^{(1)} \otimes (\mathbf{P}_i \mathbf{P}_i')$, $\mathbf{w}_i^{(P2)} \equiv \mathbf{w}_i^{(2)} \otimes \mathbf{P}_i$ and $\mathbf{w}_i^{(P3)} \equiv \mathbf{w}_i^{(3)} \otimes \mathbf{P}_i$.

Still within the framework of moment-based statistics, finally we consider the test proposed in Butler and Chatterjee (1997), also designed against a general alternative. Note that (1) implies that $E\{X_{li} m_{ji}(\theta_0)\} = 0$, for $l = 1, \dots, k$, $j = 1, \dots, J$, where X_{li} denotes the l -th component of X_i . If Jk is greater than the number of parameters (i.e. if $J \geq 2$ and $k \geq 2$), these Jk moment conditions can be used to perform a test of overidentifying restrictions. Adapting the results of the generalized

method of moments (GMM) to our framework, it follows that the test can be computed as follows: i) obtain an initial estimate of θ_0 , say $\bar{\theta}$, by minimizing $\mathbf{s}_n(\theta)' \mathbf{s}_n(\theta)$, where $\mathbf{s}_n(\theta) \equiv n^{-1} \sum_{i=1}^n \mathbf{m}_i(\theta) \otimes X_i$; ii) compute $\mathbf{S}_n(\bar{\theta})$, where $\mathbf{S}_n(\theta)$ is a $Jk \times Jk$ matrix with (j, j) submatrix $\mathbf{S}_{jj,n}(\theta) = n^{-1} \sum_{i=1}^n X_i X_i' p_{ji}(\theta) \{1 - p_{ji}(\theta)\}$, and (j, l) submatrix $\mathbf{S}_{jl,n}(\theta) = -n^{-1} \sum_{i=1}^n X_i X_i' p_{ji}(\theta) p_{li}(\theta)$ for $j \neq l$; iii) obtain a final estimate of θ_0 , say $\tilde{\theta}$, by minimizing $\mathbf{s}_n(\theta)' \mathbf{S}_n(\bar{\theta})^{-1} \mathbf{s}_n(\theta)$; and iv) compute the statistic $C_n^{(BC)} = n \mathbf{s}_n(\tilde{\theta})' \mathbf{S}_n(\bar{\theta})^{-1} \mathbf{s}_n(\tilde{\theta})$. From the results of GMM theory, it follows that if (1) holds, then $C_n^{(BC)} \xrightarrow{d} \chi_{Jk-(k+J-1)}^2$.

3. STATISTICS BASED ON COMPARISONS BETWEEN PARAMETRIC AND NONPARAMETRIC ESTIMATIONS

Many specification tests have been developed by comparing parametric and non-parametric estimations. We focus here on four of them: the ones proposed in Andrews (1997), Stute and Zhu (2002), Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005). As we discuss below, the statistic proposed in Andrews (1997) is the only one that applies directly to our problem, but the others can also be adapted to our framework. We restrict our attention to these statistics since we want to focus on statistics whose behavior does not depend crucially on the choice of a smoothing value: note that two of them use no smoothing value at all, whereas the other two partially rule out the influence of smoothing value selection by using as a statistic a maximum from among statistics computed with different smoothing values. In all cases a root- n -consistent estimator of θ_0 is required; as above, the ML estimator is the natural choice.

Andrews (1997) suggests testing a parametric specification of the conditional distribution $Y | X = x$ by comparing the joint empirical cdf of $(Y, X)'$ and an estimate of the joint cdf based on the parametric specification. Specifically, he proposes the Kolmogorov-Smirnov-type statistic $C_n^{(AN)} \equiv \max_{1 \leq j \leq n} |n^{1/2} H_n(X_j, Y_j)|$, where $H_n(x, y)$

$\equiv n^{-1} \left\{ \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y, X_i \leq x\}} - \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} F(y | X_i, \theta) \right\}$ and $F(\cdot | x, \theta_0)$ is the conditional cdf specified for $Y | X = x$. The asymptotic null distribution of $C_n^{(AN)}$ cannot be tabulated, but asymptotically valid critical values can be derived using bootstrap samples $\{(Y_{ib}^*, X_{ib}^*)'\}_{i=1}^n$ obtained as follows: $X_{ib}^* = X_i$ and Y_{ib}^* is generated with cdf $F(\cdot | X_i, \hat{\theta})$.

Stute and Zhu (2002) suggest testing the specification of a generalized linear regression model by comparing non-smoothed parametric and nonparametric estimations of the regression function. Their statistic does not directly apply to our problem, since our specification is not a generalized linear regression model if $J > 1$. However, observe that (1) holds if and only if

$$E(D_{ji} | X_i) = p_{ji}(\theta_0) \quad \text{for } j = 1, \dots, J, \quad (5)$$

where $p_{ji}(\theta_0)$ can be expressed as a function of $X_i' \beta_0$ and μ_0 . Thus, our specification is equivalent to J generalized linear regression models. Hence, we can derive a test statistic for our problem as follows: i) compute $C_{j,n}^{(SZ)}$, the Cramér-von Mises-type statistic for the j -th regression model in (5) that is obtained following Stute and Zhu (2002): $C_{j,n}^{(SZ)} \equiv n^{-2} \sum_{l=1}^n \left[\sum_{i=1}^n \mathbf{1}_{\{X_i' \hat{\beta} \leq X_l' \hat{\beta}\}} \{D_{ji} - p_{ji}(\hat{\theta})\} \right]^2$; ii) derive the overall statistic $C_n^{(SZ)} \equiv \sum_{j=1}^J C_{j,n}^{(SZ)}$. Note that this overall statistic could also be defined in a different way, e.g. $\max_{1 \leq j \leq J} C_{j,n}^{(SZ)}$ or $\sum_{j=1}^J \{C_{j,n}^{(SZ)}\}^2$. The asymptotic null distribution of $C_n^{(SZ)}$ is not known, but bootstrap critical values can be derived as above; the asymptotic validity of the bootstrap in this context can be proved with similar arguments as in Stute et al. (1998). It is worth emphasizing that $C_n^{(SZ)}$ is not affected by the curse of dimensionality, which typically appears in nonparametric estimations, since the effect of regressors is always introduced by means of the so-called single index $X_i' \hat{\beta}$. However, unlike the other three statistics discussed in this Section, we cannot ensure consistency of $C_n^{(SZ)}$ against any alternative, only against alternatives in which regressors enter through a single-index¹.

¹Following Stute (1997), a test statistic consistent against any alternative can be achieved simply

Horowitz and Spokoiny (2001) propose testing the specification of a regression model comparing smoothed nonparametric and parametric estimations of the regression function with various smoothing values. To derive a test statistic for our problem, we proceed as follows: i) compute $C_{j,n}^{(HS)}$, the Horowitz-Spokoiny statistic for the j -th regression model in (5):

$$C_{j,n}^{(HS)} \equiv \max_{h \in H_{j,n}} \frac{\sum_{l=1}^n \left[\sum_{i=1}^n \{D_{ji} - p_{ji}(\hat{\theta})\} w_{i,h}(X_l) \right]^2 - \sum_{i=1}^n a_{ii,h} \hat{\sigma}_{ji}^2}{\left\{ 2 \sum_{i=1}^n \sum_{l=1}^n a_{il,h}^2 \hat{\sigma}_{ji}^2 \hat{\sigma}_{jl}^2 \right\}^{1/2}}$$

where $w_{i,h}(x) \equiv K\{(x - X_i)/h\} / \sum_{l=1}^n K\{(x - X_l)/h\}$ is the Nadaraya-Watson weight, $K(\cdot)$ is the kernel function, h is a smoothing value, $a_{il,h} \equiv \sum_{m=1}^n w_{i,h}(X_m) w_{l,h}(X_m)$, $\hat{\sigma}_{ji}^2 \equiv p_{ji}(\hat{\theta}) \{1 - p_{ji}(\hat{\theta})\}$, $H_{j,n} \equiv \{h_{j0} a_j^l\}_{l=0}^{J_n-1}$ is a grid of smoothing values, J_n is an integer and h_{j0} , a_j are fixed values, $0 < a_j < 1$; ii) derive the overall statistic $C_n^{(HS)} \equiv \sum_{j=1}^J C_{j,n}^{(HS)}$, which detects any deviation in any of the J regression models. Note that the researcher must choose J_n , h_{j0} and a_j . As before, the asymptotic null distribution of $C_n^{(HS)}$ is not known, but critical values can be derived by bootstrap. Observe that neither the bootstrap procedure nor the conditional variance estimators $\hat{\sigma}_{ji}^2$ that we use are those proposed in Horowitz and Spokoiny (2001), since we exploit that the dependent variable is binary; in this way, a better performance is obtained.

The statistic proposed in Guerre and Lavergne (2005) is similar in spirit to the previous one but, in order to maximize power, the discrepancy between the smoothed estimations is standardized with an estimate of its conditional standard deviation computed with a possibly different smoothing value; additionally, they exploit the properties of leave-one-out estimators to simplify the estimation of asymptotic conditional variances. Specifically, the statistic that has to be used to test the specification of the j -th regression model in (5) is $C_{j,n}^{(GL)} = T_{n,\tilde{h}_j}^{(j)} / \hat{v}_{h_{j0}}^{(j)}$, where $T_{n,h}^{(j)} \equiv \frac{\sum_{i=1}^n \{D_{ji} - p_{ji}(\hat{\theta})\} w_{i,h}(X_i)}{\sqrt{\sum_{i=1}^n w_{i,h}^2(X_i) \hat{\sigma}_{ji}^2}}$ replacing $\mathbf{1}_{\{X'_i \hat{\beta} \leq X'_i \beta\}}$ by $\mathbf{1}_{\{X_i \leq X_l\}}$ in the definition; but this consistent statistic does suffer from the curse of dimensionality. We do not include it in our study since in our simulations it is always outperformed by $C_n^{(SZ)}$ when the number of non-constant regressors is greater than one.

$$\sum_{i=1}^n \sum_{l=1, l \neq i}^n \omega_{il}(h) \{D_{ji} - p_{ji}(\theta)\} \{D_{li} - p_{li}(\theta)\}, \widehat{v}_h^{(j)2} \equiv 2 \sum_{i=1}^n \sum_{l=1, l \neq i}^n \omega_{il}(h)^2 \widehat{\sigma}_{ji}^2 \widehat{\sigma}_{jl}^2,$$

$$\omega_{il}(h) \equiv K\{(X_i - X_l)/h\} / [\sum_{j=1, j \neq i}^n K\{(X_j - X_i)/h\} \sum_{j=1, j \neq l}^n K\{(X_j - X_l)/h\}]^{1/2}, \widetilde{h}_j =$$

$$\arg \max_{h \in H_{j,n}} \{T_{n,h}^{(j)} - \gamma_n \widehat{v}_{h,h_{j0}}^{(j)}\} \text{ and } \widehat{v}_{h,h_{j0}}^{(j)2} \equiv 2 \sum_{i=1}^n \sum_{l=1, l \neq i}^n \{\omega_{il}(h) - \omega_{il}(h_{j0})\}^2 \widehat{\sigma}_{ji}^2 \widehat{\sigma}_{jl}^2.$$

Note that now we must choose γ_n in addition to J_n , h_{j0} , and a_j . Once $C_{j,n}^{(GL)}$ has been computed, the overall statistic $C_n^{(GL)} \equiv \sum_{j=1}^J C_{j,n}^{(GL)}$ has to be derived. As above, critical values can be derived by bootstrap.

4. SIMULATION STUDY AND EMPIRICAL APPLICATION

We perform two sets of Monte Carlo experiments. First, we generate n i.i.d. observations from 18 different models. The dependent variable is $Y_i = \mathbf{1}_{\{Y_i^* \geq 0\}}$ in models labelled with an A and $Y_i = \mathbf{1}_{\{Y_i^* \geq 0\}} + \mathbf{1}_{\{Y_i^* \geq \mu_0\}}$ in models labelled with a B, and the latent variable Y_i^* is generated as follows:

- Models 1A, 1B: $Y_i^* = \beta_{01} + \beta_{02} X_{2i} + c(X_{2i}^2 - 1) + u_i$, where X_{2i} and u_i are independent, $X_{2i}, u_i \sim N(0, 1)$.
- Models 2A, 2B: Y_i^*, X_{2i} as in Model 1A, and $u_i | X_{2i} = x_2 \sim N(0, \exp(cx_2 - c^2/2))$.
- Models 3A, 3B: Y_i^*, X_{2i} as in Model 1A, X_{2i} and u_i independent, and if $c = 0$ then $u_i \sim N(0, 1)$; otherwise $u_i = (\mathbf{1}_{\{c > 0\}} - \mathbf{1}_{\{c < 0\}})(|c|^{1/2} \varepsilon_i + |c|^{-1/2})$, where ε_i has density function $f_\varepsilon(x) = x^{(1/|c|)-1} \exp(-x) / \Gamma(1/|c|)$ (if $c \approx 0$, u_i is approximately $N(0, 1)$; if $|c|$ is large the distribution of u_i is highly skewed).
- Models 4A, 4B, 7A, 7B: $Y_i^* = \beta_{01} + \beta_{02} X_{2i} + \beta_{03} X_{3i} + \beta_{04} X_{4i} + c(X_{2i}^2 - 1)(X_{3i}^2 - 1)(X_{4i}^2 - 1) + u_i$, where X_{2i}, X_{3i}, X_{4i} and u_i are independent with distribution $N(0, 1)$.
- Models 5A, 5B, 8A, 8B: $Y_i^*, X_{2i}, X_{3i}, X_{4i}$ as in Model 4A, and $u_i | (X_{2i} = x_2, X_{3i} = x_3, X_{4i} = x_4) \sim N(0, \exp(cx_2 - c^2/2))$
- Models 6A, 6B and 9A, 9B: $Y_i^*, X_{2i}, X_{3i}, X_{4i}$ as in Model 4A, X_{2i}, X_{3i}, X_{4i} and u_i are independent, and u_i as in Model 3A.

The true parameter values are $\beta_{01} = 0$, $\beta_{0j} = 1$ (for $j > 1$) in Models 1A-6A;

$\beta_{01} = 0$, $\beta_{0j} = 1/3$ (for $j > 1$) in Models 7A-9A; $\mu_0 = 2$, $\beta_{0j} = 1$ (for all j) in Models 1B-6B; and $\mu_0 = 2$, $\beta_{0j} = 1/3$ (for all j) in Models 7B-9B. In all models we test the null hypothesis that (1) holds with the standard normal cdf as $F(\cdot)$. Parameters are estimated by ML assuming that H_0 holds. Note that H_0 is true if and only if $c = 0$, and values of c different from 0 allow us to examine the ability of the statistics to detect misspecification in the latent regression model due to non-linearities (Models 1, 4, 7), heteroskedasticity (Models 2, 5, 8) and non-normality in $F(\cdot)$ (Models 3, 6, 9). Also note that we consider models with one non-constant regressor (Models 1-3), and with three non-constant regressors, either with a high R^2 in the latent regression model (Models 4-6) or with a low R^2 (Models 7-9). In Tables 1-9 we report the proportion of rejections of H_0 at the 5% significance level when the sample size is $n = 250$. The results are based on 1000 simulation runs, performed using GAUSS programmes that are available from the authors on request. The experiments are also run with $n = 100$ and $n = 500$, but these results are not reported since they lead to similar conclusions.

When computing $C_{j,n}^{(HS)}$ and $C_{j,n}^{(GL)}$ a grid of smoothing values $H_{j,n}$ must be chosen. We expected this choice not to be important, but the results from some preliminary samples suggest that this is not entirely the case. It is not within the scope of this paper to study the optimal choice of this grid, and we failed to locate references where hints on how to select it are given. Finally, we adopt this relatively automatic procedure: i) compute $h_{CV}^{(j)}$, the leave-one-out cross-validation bandwidth in the Nadaraya-Watson nonparametric estimation of $E[\mathbf{1}_{\{Y=j\}} \mid X = \cdot]$; ii) consider the grid $H_{j,n} = \{h_{j0} a_j^l\}_{l=0}^{J_n-1}$, where $h_{j0} = \varphi_j h_{CV}^{(j)} / a_j^{J_n-1}$, J_n is the closest integer to $\ln n$, and $a_j \in (0, 1)$, φ_j are fixed values. Note that the minimum value in $H_{j,n}$ is $O(h_{CV}^{(j)})$, and $H_{j,n}$ has the structure that is required in the theoretical assumptions. This procedure is easy to implement and only requires selecting a_j , φ_j . In our experiments, all values of a_j within $[0.75, 0.95]$ lead to almost identical results; the performance of the statistics is more sensitive to

the choice of φ_j , but in our experiments all values within $[1.5, 2]$ lead to approximately correct empirical sizes -note that this means that some “oversmoothing” is required. We finally choose $a_j = 0.9$, and $\varphi_j = 1.5$ if the model only has one non-constant regressor or $\varphi_j = 1.8$ otherwise. In the models with various non-constant regressors, the same smoothing values are used with all regressors (all of them have unit variance).

The other choices required to compute the statistics are less crucial for the results. As a kernel function we use the unit-variance density $K(u) = \mathbf{1}_{\{|u| \leq \sqrt{7}\}} \{15/(16\sqrt{7})\}(1 - u^2/7)^2$, or a product of densities of this type. When computing $C_n^{(GL)}$ we take $\gamma_{250} = \{2 \ln(J_n - 1)\}^{1/2}$. When computing $C_{n,l}^{(MP)}$ we consider various partitions of the support, but we only report the results for the statistic that performs better, namely, $C_{n,3}^{(MP)}$ with $G = 2$ in Models 1-3, and with $G = 8$ in Models 4-9 (the partitions are obtained splitting the support of each non-constant regressor in $(-\infty, 0)$ and $[0, \infty)$). When required, 101 bootstrap replications are used. Bootstrap versions of the moment-based statistics are also computed and their results are reported with an asterisk (for computational reasons, bootstrap versions of $C_n^{(BC)}$ are not computed).

First we discuss the results for the moment-based statistics. We do not report the results for $C_{n,1}^{(M)}$ and $C_{n,2}^{(M)}$ used with χ^2 critical values, since their empirical size is much higher than the nominal size (when the nominal size is 0.05 the empirical size is above 0.10 in most cases); we do report the results for $C_{n,3}^{(M)}$, which behaves well in terms of size. The bootstrap versions of the three statistics also behave well in terms of size, though a slight size distortion is observed; with additional experiments we have checked that this problem disappears if the number of bootstrap replicates and the number of simulation runs are increased. As regards power, note that $C_{n,3}^{(M)*}$ slightly outperforms $C_{n,3}^{(M)}$; and comparing the bootstrap versions of the three statistics, $C_{n,1}^{(M)*}$ yields the worst results, and $C_{n,3}^{(M)*}$ usually outperforms $C_{n,2}^{(M)*}$, though both perform similarly in most cases (but not always, see e.g. Table 7B). If we compare $C_{n,3}^{(M)}$ and

$C_{n,3}^{(MP)}$, we observe that introducing partitions leads to increases in power only in a few models and for some specific partitions (note that we only report the results for the partition with better behaviour); hence, if the aim is to derive omnibus specification tests that yield good results in as many settings as possible, the advice has to be not to include partitions. The test of overidentifying restrictions $C_n^{(BC)}$ almost always yields the worst results, its empirical size worsens as the number of regressors increases, and may be unable to detect heteroskedastic alternatives (see Table 2B). Summarizing, the preferred moment-based statistic should be $C_{n,3}^{(M)}$ used with χ^2 critical values, or its bootstrap implementation $C_{n,3}^{(M)*}$.

As regards the statistics based on the comparison between parametric and nonparametric estimations, all four behave reasonably well in terms of size, though there is a slight tendency to overreject, especially with $C_n^{(AN)}$ and $C_n^{(SZ)}$, and this problem is more severe when $n = 100$; however this size distortion may simply be a consequence of the limited number of simulations and bootstraps. When comparing power performance, first we observe that $C_n^{(AN)}$ almost always yields the worst results; moreover, though the theoretical properties of this statistic do not depend on the number of regressors, here we observe that its behaviour worsens dramatically as the number of regressors increases. It is also observed that the two bandwidth-based statistics $C_n^{(HS)}$ and $C_n^{(GL)}$ perform similarly; in some cases the latter behaves slightly better in terms of power, but this is not a general rule. With additional experiments (not included here) we observe that $C_n^{(GL)}$ is more sensitive to bandwidth selection than $C_n^{(HS)}$. In fact, it is possible to improve the power of $C_n^{(GL)}$ in the models where it performs worse by using a different grid of smoothing values; however, we report only the results obtained with the grid described above, since our aim is to compare test statistics for which the choice of smoothing value does not play a crucial role.

The comparison between the bandwidth-based statistics and $C_n^{(SZ)}$ is not straight-

forward, let alone if we also want to include $C_{n,3}^{(M)}$ in the comparison. However, some significant conclusions can be drawn from Tables 1-9. First, we observe that in the models with one non-constant regressor $C_{n,3}^{(M)}$ performs better than the others in detecting non-linearities or heteroskedasticity, but with non-normal alternatives the evidence is mixed: in the binary choice models $C_{n,3}^{(M)}$, $C_n^{(HS)}$ and $C_n^{(GL)}$ perform similarly and outperform $C_n^{(SZ)}$, but the latter performs better when $J > 1$. Second, in the models with various non-constant regressors and a high latent R^2 , $C_n^{(HS)}$ and $C_n^{(GL)}$ perform better than the others in detecting non-linearities or heteroskedasticity, but again the evidence is mixed with non-normal alternatives: in the binary choice model $C_{n,3}^{(M)}$ is the best one to capture non-normality, but if $J > 1$ $C_n^{(SZ)}$ performs better. Third, in the models with various non-constant regressors and a low latent R^2 , $C_n^{(SZ)}$ almost always outperforms $C_n^{(HS)}$ and $C_n^{(GL)}$ (the exception to this is Model 7B); but the comparison between $C_{n,3}^{(M)}$ and $C_n^{(SZ)}$ is not so clear-cut: the former performs better to detect heteroskedasticity, but there is mixed evidence with the other alternatives. Finally, note that in most of our models the statistics' rankings are similar under both non-linear and heteroskedastic alternatives; possibly, this is not a surprising result, since heteroskedasticity can be seen as a kind of specific nonlinearity.

As a summary, one could say that in models with a single explanatory variable $C_{n,3}^{(M)}$ should be preferred, since it is easily computed, requires no subjective choice, and yields the best results in almost all cases. In models with various explanatory variables it is harder to give a general advice. If the fit of the model is good, the curse of dimensionality does not seem to affect dramatically the performance of $C_n^{(HS)}$ and $C_n^{(GL)}$, which yield on average the best results; but if the fit of the model is poor no general rule is found. However, assessing the advantages and disadvantages of each statistic, it is likely that practitioners feel more comfortable using $C_{n,3}^{(M)}$ or $C_n^{(SZ)}$ rather than $C_n^{(HS)}$ or $C_n^{(GL)}$, since those have a much lower computational cost and, what is

more important, do not require the choice of any smoothing value.

To examine if the conclusions provided by the previous experiments still hold in models closer to the reality, we perform a second set of experiments in which the regressors are not simulated, but fixed at each replication and equal to real variables contained in the Panel Study of Income Dynamics (PSID). More specifically, we extract from the 2001 Wave a sample composed by 2280 women whose age is between 21 and 40, and whose marital status is other than “never married”, and consider the vector of regressors $X_i = (1, A_i, A_i^2, E_i, D_i, H_i)$, where A_i is age, E_i is education, D_i is a dummy variable for children/no children and H_i is husband’s labor income. We randomly extract a subsample of 1000 observations and with this subsample we generate two models with dependent variable $Y_i = \mathbf{1}_{\{Y_i^* \geq 0\}} + \mathbf{1}_{\{Y_i^* \geq \mu_0\}}$, where $Y_i^* = X_i' \beta_0 + u_i$ for $i = 1, \dots, 1000$, and u_i is generated independently from X_i as follows:

- Model 10: If $c = 0$ then $u_i \sim N(0, 1^2)$, and if $c \neq 0$ then $u_i = \{(c^{-1} - 2)/c^{-1}\}^{1/2} \varepsilon_i$, where ε_i follows a Student’s t distribution with c^{-1} degrees of freedom.
- Model 11: $u_i = \{(c + 1)/(c + 3)\}^{1/2} \varepsilon_i$, where ε_i follows a Student’s t distribution with $c + 3$ degrees of freedom.

In Model 10 we test the null hypothesis that (1) holds with the standard normal cdf as $F(\cdot)$, and in Model 11 we test the null hypothesis that (1) holds with the standardized Student’s t_3 cdf as $F(\cdot)$. In both models H_0 is true if and only if $c = 0$; values of $c \neq 0$ allow us to examine the ability of the statistics to detect misspecification at the tails of u . The true parameter values are the ML estimates obtained in the empirical application described below; parameters are estimated by ML assuming that H_0 holds.

In Tables 10-11 we report the proportion of rejections of H_0 at the 5% significance level. These results are based on 200 simulation runs. The statistic $C_{n,3}^{(MP)}$ is computed with $G = 2$ and partitioning the support of the regressors according to education (higher or lower than the mean level). When computing $C_n^{(HS)}$ and $C_n^{(GL)}$ we use the

same kernels as above with the continuous regressors (which are previously standardized to have unit variance) and non-smoothing weights with the binary regressor. We use the grid of smoothing values described above with $a = 0.9$ and $\varphi = 1.8$ and take $\gamma_{1000} = (2 \ln 6)^{1/2}$. When required, we use 101 bootstrap replications. The results show that the statistics perform reasonably well to detect departures from a normal cdf due to fatter tails (Model 10), but departures from a Student's t_3 cdf due to lower tails (Model 11) are much harder to capture. As for the relative performance of the statistics, the conclusions drawn from the first set of experiments are confirmed.

As an empirical application we consider the determinants of women's labour market participation and the type of participation (full-time or part-time). We use the 2280 observations described above, and also consider the variable HW = hours of work per year. Then we define the dependent variable $Y = 0$ if $HW = 0$, $Y = 1$ if $0 < HW < 1440$, or $Y = 2$ if $HW \geq 1440$. Using the vector of regressors X_i described above we estimate an ordered probit model and an ordered discrete choice model with a standardized Student's t_3 cdf as $F(\cdot)$. In both cases all estimated coefficients have the expected sign, but the significance of some of the regressors depends on the model that is being estimated: husband's income is only significant in the ordered probit model, whereas age^2 is only significant in the model with a standardized Student's t_3 cdf. However, the ordered probit specification is rejected with all the statistics (the p-values range from 0.001 to 0.025), but the Student's t_3 specification is almost never rejected at the usual significance levels (most p-values are above 0.10).

5. CONCLUDING REMARKS

We discuss how to test the specification of an ordered discrete choice model against a general alternative. On the one hand, we consider various moment-based statistics; they are all based on a finite set of unconditional moment restrictions and, hence,

are not consistent against any possible alternative. On the other hand, we consider statistics based on comparisons between parametric and nonparametric estimations which are either consistent against any alternative (Andrew's statistic, Horowitz and Spokoiny's statistic and Guerre and Lavergne's statistic) or consistent against alternatives in which regressors enter through a single index (Stute and Zhu's statistic). We analyze how all these statistics can be implemented in our context and compare their relative performance with simulations that allow us to examine their ability to detect non-linearities, heteroskedasticity or non-normality in the latent regression model.

Moment-based statistics are easy to derive and, as usual, their finite-sample performance depends on the covariance-matrix estimator that is used. We propose a variant of the standard moment-based statistics that uses a conditional-expectation-based covariance matrix, and prove that this variant can be computed using an artificial regression. The resulting statistic performs very well both in terms of size and power, particularly in models with a single explanatory variable. It must also be stressed that in most of our simulations this moment-based statistic outperforms the widely-used test of overidentifying restrictions proposed in Butler and Chatterjee (1997). Andrew's statistic also performs reasonably well in models with a single explanatory variable, but its behaviour is severely affected by the number of regressors. The generalizations of the statistics in Stute and Zhu (2002), Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005) that we consider here also yield satisfactory results, though the latter ones require the choice of a grid of smoothing values, and their behaviour is more sensitive to this choice than we expected to. The typical curse of dimensionality of nonparametric estimations affects the performance of Horowitz and Spokoiny's and Guerre and Lavergne's statistics more dramatically in models with poor fit. However, this is not entirely the case with Stute and Zhu's statistic, which exhibits good power properties in all models with various regressors, possibly because it circumvents the

course of dimensionality by the use of a single index.

Finally, if we try to compare the performance of all the statistics, we observe that in models with only one non-constant regressor, our variant of the moment-based statistic yields the best results. But when the number of non-constant regressors is greater than one, the results are less conclusive, since the relative performance of the statistics depends on both the fit the model and the type of misspecification that is considered.

APPENDIX: ANALYTICAL EXPRESSIONS

Hereafter, $f(\cdot)$ and $\dot{f}(\cdot)$ denote the first and second derivative of $F(\cdot)$, $f_{0i} \equiv f(-X'_i\beta)$, $\dot{f}_{0i} \equiv \dot{f}(-X'_i\beta)$, $p_{ji} \equiv p_{ji}(\theta)$, and, for $j = 1, \dots, J-1$, $f_{ji} \equiv f(\mu_j - X'_i\beta)$, $\dot{f}_{ji} \equiv \dot{f}(\mu_j - X'_i\beta)$. Additionally, $f_{-1,i} \equiv 0$, $\dot{f}_{-1,i} \equiv 0$, $f_{Ji} \equiv 0$, $\dot{f}_{Ji} \equiv 0$. To compute $\mathbf{V}_{n,1}$ and $\mathbf{V}_{n,2}$, expressions for $\mathbf{B}_i(\theta)$, $\mathbf{g}_i(\theta)$ and $\mathbf{A}_i(\theta)$ are required. From the definitions in Section 2, it follows that $\mathbf{B}_i(\theta) = [\mathbf{B}_{1i}(\theta) : \mathbf{B}_{2i}(\theta)]$, where $\mathbf{B}_{1i}(\theta)$ is the $J \times k$ matrix whose j -th row is $(f_{ji} - f_{j-1,i})X'_i$, and $\mathbf{B}_{2i}(\theta)$ is the $J \times (J-1)$ matrix whose (j, l) element is $-\mathbf{1}_{\{l=j\}}f_{ji} + \mathbf{1}_{\{l=j-1\}}f_{j-1,i}$. On the other hand, $\mathbf{g}_i(\theta)$ is the $(k + J - 1) \times 1$ vector whose first k rows are $-\{\sum_{j=0}^J (f_{ji} - f_{j-1,i})D_{ji}/p_{ji}\}X_i$, and whose $(k + l)$ -th row is $\{D_{li}/p_{li} - D_{l+1,i}/p_{l+1,i}\}f_{li}$. Finally, $\mathbf{A}_i(\theta)$ is the $k + (J-1) \times k + (J-1)$ symmetric matrix whose (1, 1) submatrix is $\{\sum_{j=0}^J D_{ji}[(f_{ji} - f_{j-1,i})^2 - p_{ji}(\dot{f}_{ji} - \dot{f}_{j-1,i})]/p_{ji}^2\}X_iX'_i$, its (1, 2) submatrix is the $k \times (J-1)$ matrix whose j -th column is $\{D_{ji}[\dot{f}_{ji}p_{ji} - (f_{ji} - f_{j-1,i})\dot{f}_{ji}]/p_{ji}^2 + D_{j+1,i}[(f_{j+1,i} - f_{ji})\dot{f}_{ji} - \dot{f}_{ji}p_{j+1,i}]/p_{j+1,i}^2\}X_i$, and its (2, 2) submatrix is the symmetric matrix whose $(j, j+l)$ element is $\mathbf{1}_{\{l=0\}}\{D_{ji}(f_{ji}^2 - \dot{f}_{ji}p_{ji})/p_{ji}^2 + D_{j+1,i}(f_{ji}^2 + \dot{f}_{ji}p_{j+1,i})/p_{j+1,i}^2\} - \mathbf{1}_{\{l=1\}}D_{j+1,i}\dot{f}_{ji}f_{j+1,i}/p_{j+1,i}^2$. To compute $\mathbf{V}_{n,3}$, expressions for $E_X\{\mathbf{m}_i(\theta)\mathbf{m}_i(\theta)'\}$, $E_X\{\mathbf{m}_i(\theta)\mathbf{g}_i(\theta)'\}$ and $E_X\{\mathbf{g}_i(\theta)\mathbf{g}_i(\theta)'\}$ are required. Here $E_X\{\mathbf{m}_i(\theta)\mathbf{m}_i(\theta)'\}$ is the $J \times J$ symmetric matrix whose (j, j) element is $p_{ji}(1 - p_{ji})$ and whose (j, l) element, for $l > j$, is $-p_{ji}p_{li}$, $E_X\{\mathbf{m}_i(\theta)\mathbf{g}_i(\theta)'\} = -\mathbf{B}_i(\theta)$, and $E_X\{\mathbf{g}_i(\theta)\mathbf{g}_i(\theta)'\}$ is the $k + (J-1) \times k + (J-1)$ symmetric matrix whose (1, 1) submatrix

is $\{\sum_{j=0}^J (f_{ji} - f_{j-1,i})^2 / p_{ji}\} X_i X_i'$, its (1, 2) submatrix is the $k \times (J - 1)$ matrix whose j -th column is $\{(f_{j+1,i} - f_{ji})/p_{j+1,i} - (f_{ji} - f_{j-1,i})/p_{ji}\} f_{ji} X_i$, and its (2, 2) submatrix is the symmetric matrix whose (j, l) element, for $l \geq j$, is $\mathbf{1}_{\{l=j\}}(1/p_{ji} + 1/p_{j+1,i})f_{ji}^2 - \mathbf{1}_{\{l=j+1\}}(f_{ji}f_{j+1,i}/p_{j+1,i})$. Finally, $\mathbf{V}_{n,3}^{(P)}$ can be computed using that $E_X\{\mathbf{m}_i^{(P)}(\theta)\mathbf{m}_i^{(P)}(\theta)'\} = E_X\{\mathbf{m}_i(\theta)\mathbf{m}_i(\theta)'\} \otimes (\mathbf{P}_i \mathbf{P}_i')$ and $E_X\{\mathbf{m}_i^{(P)}(\theta)\mathbf{g}_i(\theta)'\} = -\mathbf{B}_i^{(P)}(\theta)$.

REFERENCES

- Andrews, D. W. (1987). “Asymptotic Results for Generalized Wald Tests”, *Econometric Theory*, Vol. 3, pp. 348-358.
- Andrews, D. W. (1988). “Chi-Square Diagnostic Tests for Econometric Models”, *Journal of Econometrics*, Vol. 37, pp. 135-156.
- Andrews, D. W. (1997). “A Conditional Kolmogorov Test”, *Econometrica*, Vol. 65, pp. 1097-1128.
- Butler, J. S. and Chatterjee, P. (1997). “Tests of the Specification of Univariate and Bivariate Ordered Probit”, *Review of Economics and Statistics*, Vol. 79, pp. 343-347.
- Guerre, E. and Lavergne, P. (2005). “Data-Driven Rate-Optimal Specification Testing in Regression Models”, *Annals of Statistics*, Vol. 33, pp. 840-870.
- Horowitz, J. L. and Spokoiny, V. G. (2001). “An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative”, *Econometrica*, Vol. 69, pp. 599-631.
- Johnson, P. A. (1996). “A Test of the Normality Assumption in the Ordered Probit Model”, *Metron*, Vol. 54, pp. 213-221.
- Murphy, A. (1996). “Simple LM Tests of Mis-Specification for Ordered Logit Models”, *Economics Letters*, Vol. 52, pp. 137-141.
- Newey, W. K. (1985). “Maximum Likelihood Specification Testing and Conditional

Moment Tests”, *Econometrica*, Vol. 53, pp. 1047-1070.

Orme, C. D. (1990). “The Small Sample Performance of the Information Matrix Test”, *Journal of Econometrics*, Vol. 46, pp. 309-331.

Robin, J. M. and Smith R. J. (2000). “Tests of Rank”, *Econometric Theory*, Vol. 16, pp. 151-175.

Skeels, C. L. and Vella, F. (1999). “A Monte Carlo Investigation of the Sampling Behavior of Conditional Moment Tests in Tobit and Probit Models”, *Journal of Econometrics*, Vol. 92, pp. 275-294.

Stute, W. (1997). “Nonparametric Model Checks for Regression”, *Annals of Statistics*, Vol. 25, pp. 613-641.

Stute, W., González-Manteiga, W. and Presedo-Quindimil, M. (1998). “Bootstrap Approximations in Model Checks for Regression”, *Journal of the American Statistical Association*, Vol. 93, pp. 141-149.

Stute, W. and Zhu, L. X. (2002). “Model Checks for Generalized Linear Models”, *Scandinavian Journal of Statistics*, Vol. 29, pp. 535-545.

Tauchen, G. E. (1985). “Diagnostic Testing and Evaluation of Maximum Likelihood Models”, *Journal of Econometrics*, Vol. 30, pp. 415-443.

Weiss, A. A. (1997). “Specification Tests in Ordered Logit and Probit Models”, *Econometric Reviews*, Vol. 16, pp. 361-391.

Wooldridge, J. M. (1990). “A Unified Approach to Robust, Regression-Based Specification Tests”, *Econometric Theory*, Vol. 6, pp. 17-43.