A discusión

CONSISTENT SPECIFICATION TESTS FOR ORDERED DISCRETE CHOICE MODELS[*]

Juan Mora and Ana I. Moro-Egido[**]

WP-AD 2006-17

Correspondence to: Juan Mora. Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Apartado de Correos 99, 03080 Alicante, Spain. E-mail address: juan@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Julio 2006
Depósito Legal: V-3367-2006

[*] We are grateful to M.D. Collado, C. Muller and F. Peñaranda for helpful comments. Financial support from Spanish MEC (SEJ2005-02829/ECON) and Ivie is gratefully acknowledged.

[**]J. Mora: Universidad de Alicante. A.I. Moro-Egido: Universidad de Granada.

CONSISTENT SPECIFICATION TESTS FOR ORDERED DISCRETE CHOICE MODELS

Juan Mora and Ana I. Moro-Egido

ABSTRACT

We discuss how to test consistently the specification of an ordered discrete choice model. Two approaches are considered: tests based on conditional moment restrictions 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. A variant of the standard conditional moment statistic and a generalization of Horowitz-Spokoiny's statistic perform best.

Codes JEL: C25, C52, C15.

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

1. INTRODUCTION

Ordered discrete choice variables often appear in Statistics and Econometrics as a dependent variable. The outcomes of an ordered discrete choice variable Y are usually labelled as $0, 1, ..., J$. Given certain explanatory variables $X = (X_1, ..., X_k)'$, the researcher is usually interested in analysing whether one (or some) of the proposed explanatory variables is significant or not, and/or providing accurate estimates of the conditional probabilities $\Pr(Y = j \mid X = x)$, which may be interesting by themselves or required in a first stage to derive a two-stage estimator. Examples of ordered discrete choice dependent variables that have been used in applied work include: education level attained by individuals (Jiménez and Kugler 1987); female labour participation: work full-time, work part-time, not to work (Gustaffson and Stafford 1992); level of demand for a new product or service (Klein and Sherman 1997); and number of children in a family (Calhoun 1989).

The parametric model that is most frequently used 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:

$$Y = j \qquad \text{if} \quad \mu_{0,j-1} \leq Y^* < \mu_{0j}, \quad \text{for } j = 0, 1, ..., J, \tag{1}$$

where $\mu_{0,-1} \equiv -\infty$, $\mu_{0,J} \equiv +\infty$ and μ_{00}, μ_{01}, ..., $\mu_{0,J-1}$ are threshold parameters such that $\mu_{00} \leq \mu_{01} \leq ... \leq \mu_{0,J-1}$. Assuming independence between u and X, the relationship (1) induces the following specification for Y:

$$\Pr(Y = j \mid X) = F(\mu_{0j} - X'\beta_0) - F(\mu_{0,j-1} - X'\beta_0), \quad \text{for } j = 0, 1, ..., J, \tag{2}$$

where $F(\cdot)$ is the distribution function of u, usually referred to as the "link function". In a parametric framework, for the sake of identification of the model it is usually assumed that the first threshold parameter μ_{00} is zero; additionally it is assumed that $F(\cdot)$ is entirely known, the most typical choices being the standard normal distribution ("ordered probit model") and the logistic distribution ("ordered logit model"). With these assumptions we obtain a full parameterization of the conditional distribution $Y \mid X = x$, with parameter vector $\theta_0 = (\beta_0', \mu_0')' \in \Theta \subset \mathbb{R}^{k+J-1}$.

The key assumptions in a parametric ordered discrete choice model are the linearity in the latent regression model, the form of the link function $F(\cdot)$ (specifically, its symmetry and its behaviour at the tails) and the independence between u and X in the latent regression model (which, in turn, implies that it is homoskedastic). Parameter estimates and predicted probabilities based on ordered discrete choice models are inconsistent if any of these key assumptions is not met (note that in contrast to standard regression models, heteroskedasticity or misspecification of the distribution function of u provoke inconsistencies). On the other hand, there exist semiparametric methods (Klein and Sherman 2002) and even nonparametric methods (Matzkin 1992) that allow for consistent estimation of the conditional probabilities $\Pr(Y = j \mid X = x)$ under much weaker assumptions; these methods have not been much used in empirical applications due their technical complexity, but they provide a reasonable alternative to purely parametric methods. Therefore, it is especially important to test the specification of a parametric ordered discrete choice model, since misspecification errors lead to inconsistent estimation, and it is possible to use alternative consistent techniques.

In recent years, various statistics have been proposed to test one (or some) of the assumptions of a parametric ordered discrete choice model. However, most of these statistics are constructed to detect only specific departures from the null hypothesis; for example, Weiss (1997) proposes to test the null of a homoskedastic u versus some parametric heteroskedastic alternatives; Glewwe (1997) proposes to test the null that u is normal versus the alternative that it is a member of the Pearson family; Murphy (1996) proposes to test the null that $F(\cdot)$ is logistic versus various alternatives; and Santos Silva (2001) proposes a test statistic to face two non-nested parametric specifications. By contrast, we focus here on consistent specification test statistics, i.e. test statistics that allow us to detect any deviation from the null hypothesis "the proposed parametric specification is correct".

Roughly speaking, two main approaches may be followed in our context to derive consistent specification statistics: tests based on conditional moment (CM) restrictions, which can be constructed following the general methodology described in Newey (1985), Tauchen (1985) and Andrews (1988); and tests based on comparisons between parametric and nonparametric estimations, such as those proposed by Andrews (1997), Stute (1997) and Horowitz and Spokoiny (2001), among others. The objective of this article is to examine how these two approaches can be applied to test the specification of an ordered discrete choice model, and to compare the performance of the resulting statistics by means of simulations.

Our results highlight the importance of covariance matrix estimation in CM tests. Specifically, standard CM statistics, computed with covariance matrix estimators based on actual derivatives, usually perform worse than statistics based on comparisons between parametric and nonparametric estimations. However, exploiting the information about conditional expectations contained in the model, here we derive variants of standard CM statistics that perform much better than standard CM statistics; additionally, these variants are easy to compute, since they can be obtained using artificial regressions. On the other hand, our results suggest that the methodology proposed in Horowitz and Spokoiny (2001) can be extended successfully to the context of ordered discrete choice models, and the resulting statistic outperforms other statistics that are based on comparisons between parametric and nonparametric methods. However, when we compare the performace of the generalization of Horowitz-Spokoiny's statistic with the variant of the standard CM statistic that we propose to use, there is no clear-cut answer to the question of which one performs better: the latter outperforms the former in some cases (e.g. with heteroskedastic alternatives) and is computationally much less demanding, but the additional power that is obtained when using the former in some other cases (e.g. with non-normal alternatives) may well justify the extra computing effort.

The rest of the paper is organized as follows. In Sections 2 and 3 we describe the test statistics considered here. In Section 4 we report the results of various Monte Carlo experiments. In Section 5 we present two empirical applications. In Section 6 we conclude. Some technical details are relegated to an Appendix.

2. STATISTICS BASED ON CONDITIONAL MOMENT RESTRICTIONS

We assume that independent and identically distributed (i.i.d.) observations $(Y_i, X_i')'$ are available, where, hereafter, $i = 1, ..., n$. Additionally, the following notation will be used: $D_{ji} \equiv I(Y_i = j)$, for $j = 0, 1, ..., J$, where $I(\cdot)$ is the indicator function; 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 $j = 2, ..., J-1$.

Let us consider $m_{ji}(\theta) \equiv D_{ji} - p_{ji}(\theta)$. Then, it follows from (2) that

$$E\{m_{ji}(\theta_0) \mid X_i\} = 0, \quad \text{for } j = 0, 1, ..., J. \tag{3}$$

This yields $J + 1$ conditional moment restrictions that must be satisfied. In fact, since the sum of all probabilities adds to one, only J moment conditions are relevant to construct a test statistic. Specifically, we disregard the moment condition for $j = 0$ and consider the random vector $\sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta})$, where $\mathbf{m}_i(\theta)$ is the $J \times 1$ column vector whose j-th component is $m_{ji}(\theta)$ and $\widehat{\theta}$ is a well-behaved estimator of θ_0. To derive an asymptotically valid test statistic, we must analyze the asymptotic behaviour of $\sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta})$. First of all note that using a first-order Taylor expansion of $m_{ji}(\widehat{\theta}) - m_{ji}(\theta_0)$, it follows that

$$n^{-1/2} \sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta}) = n^{-1/2} \sum_{i=1}^{n} \mathbf{m}_i(\theta_0) + \mathbf{B}_0 \times n^{1/2}(\widehat{\theta} - \theta_0) + o_p(1), \tag{4}$$

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 $\partial m_{ji}(\theta)/\partial\theta'$. Thus, the asymptotic behaviour of $n^{-1/2} \sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta})$ depends on the asymptotic behaviour of $n^{1/2}(\widehat{\theta} - \theta_0)$. In our context, since our null hypothesis specifies the conditional distribution $Y_i \mid X_i = x$, the natural way to estimate θ_0 is maximum likelihood (ML). The log-likelihood of the model can be written as

$$\ln L(\theta) = \sum_{i=1}^{n} \sum_{j=0}^{J} D_{ji} \ln p_{ji}(\theta).$$

We assume that the regularity conditions that ensure that the ML estimator is asymptotically normal are met (see e.g. Amemiya 1985, Chapter 9). In this case, this amounts to saying that the ML estimator $\widehat{\theta}$ satisfies

$$n^{1/2}(\widehat{\theta} - \theta_0) = \mathbf{A}_0^{-1} \times n^{-1/2} \sum_{i=1}^{n} \mathbf{g}_i(\theta_0) + o_p(1), \tag{5}$$

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 $\mathbf{A}_0$ is the limiting information matrix, i.e. $\mathbf{A}_0 = E\{\mathbf{A}_i(\theta_0)\}$, for $\mathbf{A}_i(\theta) \equiv -\partial\mathbf{g}_i(\theta)/\partial\theta'$. From (4) and (5) it follows that:

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{m}_i(\widehat{\theta}) = [\mathbf{I}_J : \mathbf{B}_0\mathbf{A}_0^{-1}]\begin{bmatrix} n^{-1/2}\sum_{i=1}^{n}\mathbf{m}_i(\theta_0) \\ n^{-1/2}\sum_{i=1}^{n}\mathbf{g}_i(\theta_0) \end{bmatrix} + o_p(1),$$

where $\mathbf{I}_J$ is the $J \times J$ identity matrix. Hence,

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{m}_i(\widehat{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_0), \tag{6}$$

where

$$\mathbf{V}_0 \equiv [\mathbf{I}_J : \mathbf{B}_0\mathbf{A}_0^{-1}]\mathbf{Q}_0[\mathbf{I}_J : \mathbf{B}_0\mathbf{A}_0^{-1}]', \tag{7}$$

$\mathbf{Q}_0 \equiv E\{\mathbf{Q}_i(\theta_0)\}$ and $\mathbf{Q}_i(\theta) \equiv (\mathbf{m}_i(\theta)', \mathbf{g}_i(\theta)')'(\mathbf{m}_i(\theta)', \mathbf{g}_i(\theta)')$. Finally, to derive a test statistic, a consistent estimator of $\mathbf{V}_0$ must be proposed. It is worthwhile discussing in detail how $\mathbf{V}_0$ can be estimated, since it is well-known that the finite-sample performance of statistics based on conditional moment restrictions crucially depends on this.

According to (7), 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 $\mathbf{Q}_n \equiv n^{-1}\sum_{i=1}^{n}\mathbf{Q}_i(\widehat{\theta})$, $\mathbf{B}_n \equiv n^{-1}\sum_{i=1}^{n}\mathbf{B}_i(\widehat{\theta})$ and $\mathbf{A}_n \equiv n^{-1}\sum_{i=1}^{n}\mathbf{A}_i(\widehat{\theta})$. However, following the literature on artificial regressions (see e.g. MacKinnon 1992), it is possible to propose an alternative estimator of $\mathbf{V}_0$ that leads to a computationally simpler statistic. To derive it, note that $E\{m_{li}(\theta)\mathbf{g}_i(\theta)'\} = E\{p_{li}\partial \ln p_{li}(\theta)/\partial\theta'\} = E\{-\partial m_{li}(\theta)/\partial\theta'\}$; hence $E\{\mathbf{m}_i(\theta_0)\mathbf{g}_i(\theta_0)'\} = -\mathbf{B}_0$. Moreover, the information matrix equality ensures that $E\{\mathbf{g}_i(\theta_0)\mathbf{g}_i(\theta_0)'\} = \mathbf{A}_0$. Using these equalities, it follows that $\mathbf{V}_0$ equals

$$E\{\mathbf{m}_i(\theta_0)\mathbf{m}_i(\theta_0)'\} - E\{\mathbf{m}_i(\theta_0)\mathbf{g}_i(\theta_0)'\}E\{\mathbf{g}_i(\theta_0)\mathbf{g}_i(\theta_0)'\}^{-1}E\{\mathbf{g}_i(\theta_0)\mathbf{m}_i(\theta_0)'\}. \tag{8}$$

This leads us to consider

$$\mathbf{V}_{n,2} \equiv n^{-1}[\sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta})\mathbf{m}_i(\widehat{\theta})' - \sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta})\mathbf{g}_i(\widehat{\theta})'\{\sum_{i=1}^{n} \mathbf{g}_i(\widehat{\theta})\mathbf{g}_i(\widehat{\theta})'\}^{-1} \sum_{i=1}^{n} \mathbf{g}_i(\widehat{\theta})\mathbf{m}_i(\widehat{\theta})'].$$

$\mathbf{V}_{n,1}$ and $\mathbf{V}_{n,2}$ are the standard choices for estimating $\mathbf{V}_0$. Both are obtained by simply replacing population moments by sample moments. However, in our context we can do better than that. Observe that our null hypothesis yields the specification of the conditional distribution $Y \mid X = x$. This means that we can compute the conditional expectation with respect to the independent variables and then, by the law of iterated expectations, the sample analog of this conditional expectation is a consistent estimator of the the population moment; e.g. $E\{\mathbf{m}_i(\theta_0)\mathbf{m}_i(\theta_0)'\}$ can be consistently estimated with $n^{-1}\sum_{i=1}^{n} E_X\{\mathbf{m}_i(\widehat{\theta})\mathbf{m}_i(\widehat{\theta})'\}$. Following this approach, expression (8) for $\mathbf{V}_0$ suggests that we can estimate this matrix with $\mathbf{V}_{n,3} \equiv n^{-1}\sum_{i=1}^{n} \mathbf{V}_{i,3}(\widehat{\theta})$, where

$$\mathbf{V}_{i,3}(\theta) \equiv E_X\{\mathbf{m}_i(\theta)\mathbf{m}_i(\theta)'\} - E_X\{\mathbf{m}_i(\theta)\mathbf{g}_i(\theta)'\}E_X\{\mathbf{g}_i(\theta)\mathbf{g}_i(\theta)'\}^{-1}E_X\{\mathbf{g}_i(\theta)\mathbf{m}_i(\theta)'\}.$$

Observe that the analytical expressions of these conditional expectations are easy to derive. On the other hand, this approach could also be followed using expression (7) rather than (8) as a starting point; but the estimator that is obtained in this way proves to be again $\mathbf{V}_{n,3}$.

To sum up, we can consider three possible consistent estimates for $\mathbf{V}_0$ and, thus, we can derive three possible test statistics

$$C_{n,l}^{(M)} \equiv n^{-1}\{\sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta})'\}\{\mathbf{V}_{n,l}\}^{-1}\{\sum_{i=1}^{n} \mathbf{m}_i(\widehat{\theta})\},$$

for $l = 1, 2, 3$. From (6), it follows that if specification (2) is correct then $C_{n,l}^{(M)} \xrightarrow{d} \chi^2_J$; hence, given a significance level α, an asymptotically valid critical region is

$\{C_{n,l}^{(M)} > \chi^2_{1-\alpha;J}\}$, where $\chi^2_{1-\alpha;J}$ is the $1-\alpha$ quantile of a χ^2_J distribution. To facilitate the computation of these three statistics, in Appendix A1 we derive the specific analytical expressions for $\mathbf{V}_{n,1}$, $\mathbf{V}_{n,2}$ and $\mathbf{V}_{n,3}$ that are obtained here. As is well-known in the relevant literature, note that $C_{n,2}^{(M)}$ can also be computed using an artificial regression because, since $\sum_{i=1}^n \mathbf{g}_i(\widehat{\theta}) = (\partial \ln L/\partial\theta)|_{\theta=\widehat{\theta}} = 0$, the statistic $C_{n,2}^{(M)}$ proves to be the explained sum of squares in the artificial regression of a vector of ones on $\mathbf{m}_i(\widehat{\theta})'$ and $\mathbf{g}_i(\widehat{\theta})$. On the other hand, using a Cholesky decomposition of $E_X\{\mathbf{m}_i(\theta)\mathbf{m}_i(\theta)'\}$, it is also possible to derive an artificial regression whose explained sum of squares coincides with $C_{n,3}^{(M)}$ (see Appendix A2).

Still within the framework of tests based on conditional moment restrictions, more statistics can be derived following the methodology described in Andrews (1988), who proposes increasing the degrees of freedom of the statistic by partitioning the support of the regressors. Specifically, let us assume that the support of X_i is partitioned into G subsets $A_1, ..., A_G$. Then we can define $m_{jgi}(\theta) \equiv m_{ji}(\theta)I(X_i \in A_g)$ for $j = 1, ..., J$ and $g = 1, ..., G$, and thus consider the JG conditional moment restrictions

$$E\{m_{ji}(\theta_0) \mid X_i\} = 0, \quad \text{for } j = 1, ..., J, \text{ and } g = 1, ..., G.$$

To derive a test statistic, we consider now the random vector $\sum_{i=1}^n \mathbf{m}_i^{(P)}(\widehat{\theta})$, where $\mathbf{m}_i^{(P)}(\theta) \equiv \mathbf{m}_i(\theta) \otimes \mathbf{P}_i$, and $\mathbf{P}_i$ is the $G \times 1$ matrix whose g-th row is $I(X_i \in A_g)$. As before, it follows that

$$n^{-1/2} \sum_{i=1}^n \mathbf{m}_i^{(P)}(\widehat{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_0^{(P)}), \tag{9}$$

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 candidate for estimating $\mathbf{V}_0^{(P)}$ 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 $\mathbf{B}_n^{(P)} \equiv n^{-1}\sum_{i=1}^n \mathbf{B}_i^{(P)}(\widehat{\theta})$ and $\mathbf{Q}_n^{(P)} \equiv n^{-1}\sum_{i=1}^n \mathbf{Q}_i^{(P)}(\widehat{\theta})$. With the same reasoning as before, two other consistent estimators of $\mathbf{V}_0^{(P)}$ 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 $\mathbf{m}_i(\widehat{\theta})$ by $\mathbf{m}_i^{(P)}(\widehat{\theta})$ (see Appendix A1). In this fashion we again obtain three different test statistics

$$C_{n,l}^{(MP)} \equiv n^{-1}\{\sum_{i=1}^n \mathbf{m}_i^{(P)}(\widehat{\theta})'\}\{\mathbf{V}_{n,l}^{(P)}\}^{-1}\{\sum_{i=1}^n \mathbf{m}_i^{(P)}(\widehat{\theta})\},$$

for $l = 1, 2, 3$. From (9), it follows that if specification (2) is correct then $C_{n,l}^{(MP)} \xrightarrow{d} \chi^2_{JG}$; hence, $C_{n,l}^{(MP)}$ can also be used as a test statistic. Additionally, both $C_{n,2}^{(MP)}$ and $C_{n,3}^{(MP)}$ can be computed using an artificial regression. Specifically, $C_{n,2}^{(MP)}$ coincides with the explained sum of squares in the artificial regression of a vector of ones on $\mathbf{m}_i^{(P)}(\widehat{\theta})$ and $\mathbf{g}_i(\widehat{\theta})$, whereas the artificial regression which allows us to compute $C_{n,3}^{(MP)}$ is described in Appendix A2.

Finally, when $J \geq 2$ and $k \geq 2$, Butler and Chatterjee (1997) propose using a test of overidentifying restrictions. Observe that (3) implies that the following Jk moment conditions hold:

$$E\{X_{li}m_{ji}(\theta_0)\} = 0, \quad \text{for } l = 1, ..., k, \; j = 1, ..., J,$$

where X_{li} denotes the l-th component of X_i. Since the number of parameters is $k + J - 1$, a test of overidentifying restrictions is possible if $Jk > k + J - 1$, and this condition holds if and only if $J \geq 2$ and $k \geq 2$. Adapting the general results of the generalized method of moments tests to our framework (see e.g. Hamilton 1994, Chapter 14), it follows that now the test of overidentifying restrictions can be computed as follows: i) obtain an initial estimate of θ_0, say $\overline{\theta}$, 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 $\mathbf{S}_n(\overline{\theta})$, where

$$\mathbf{S}_n(\theta) \equiv \begin{bmatrix} \mathbf{S}_{11,n}(\theta) & ... & \mathbf{S}_{1J,n}(\theta) \\ ... & ... & ... \\ \mathbf{S}_{J1,n}(\theta) & ... & \mathbf{S}_{JJ,n}(\theta) \end{bmatrix},$$

$\mathbf{S}_{jj,n}(\theta) = n^{-1}\sum_{i=1}^n X_i X_i' p_{ji}(1-p_{ji})$ and $\mathbf{S}_{jl,n}(\theta) = -n^{-1}\sum_{i=1}^n X_i X_i' p_{ji} p_{li}$ for $j \neq l$; iii) obtain a final estimate of θ_0, say $\widetilde{\theta}$, by minimizing $\mathbf{s}_n(\theta)'\mathbf{S}_n(\overline{\theta})^{-1}\mathbf{s}_n(\theta)$; and iv) compute the test statistic

$$C_n^{(BC)} = n\mathbf{s}_n(\widetilde{\theta})'\mathbf{S}_n(\overline{\theta})^{-1}\mathbf{s}_n(\widetilde{\theta}).$$

If specification (2) is correct then $C_n^{(BC)} \xrightarrow{d} \chi^2_{Jk-(k+J-1)}$; thus, $C_n^{(BC)}$ can also be used as a statistic to perform an asymptotically valid specification test.

3. STATISTICS BASED ON COMPARISONS BETWEEN PARAMETRIC AND NONPARAMETRIC ESTIMATIONS

Many specification tests have recently been developed using comparisons between parametric and nonparametric estimations. In this paper we consider three of them: one based on the comparison of joint empirical distribution functions (Andrews 1997), and two others based on comparisons between regression estimations, either non-smoothed (Stute 1997) or smoothed (Horowitz and Spokoiny 2001). In fact, as we discuss below, only the first of these statistics applies directly to our framework, but the other two can be conveniently modified to cover our problem. We focus on these three statistics because they have the advantage that their performance does not depend on the choice of a bandwidth value[1]. Note as

[1] The test statistics proposed by Andrews (1997) and Stute (1997) do not use any bandwidth. Horowitz and Spokoiny (2001) propose using as a test statistic a maximum from among statistics computed with different bandwidths; hence the influence of bandwidth selection is ruled out.

well that these three statistics require the use of a root-n-consistent estimator of θ_0; as in the previous section, the ML estimator is the natural choice.

Andrews (1997) suggests testing a parametric specification of the conditional distribution $Y \mid X = x$ by comparing the joint empirical distribution function of $(Y, X')'$ and an estimate of the joint distribution function based on the parametric specification. Specifically, if $F(\cdot \mid x, \theta_0)$ is the assumed parametric conditional distribution function of $Y \mid X = x$ and we denote

$$H_n(x,y) \equiv n^{-1}\{\sum_{i=1}^{n} I(Y_i \leq y, X_i \leq x) - \sum_{i=1}^{n} F(y \mid X_i, \widehat{\theta}) I\left(X_i \leq x\right)\},$$

where $\widehat{\theta}$ is a root-n-consistent estimator of θ_0, Andrews (1997) proposes using the Kolmogorov-Smirnov-type test statistic:

$$C_n^{(AN)} \equiv \max_{1 \leq j \leq n} \left| n^{1/2} H_n(X_j, Y_j) \right| .$$

The asymptotic null distribution of $C_n^{(AN)}$ cannot be tabulated but, given a significance level α, an asymptotically valid critical region is $\{C_n^{(AN)} > c_{\alpha,n}^{(AN)}\}$, where $c_{\alpha,n}^{(AN)}$ is a bootstrap critical value. This bootstrap critical value can be obtained as the $1-\alpha$ quantile of $\{C_{n,b}^{(AN)*},\ b = 1, ..., B\}$, where $C_{n,b}^{(AN)*}$ is a bootstrap statistic constructed in the same way as $C_n^{(AN)}$, but using as sample data the b-th bootstrap sample $\{(Y_{ib}^*, X_{ib}^{*\prime})'\}_{i=1}^{n}$, which, in turn, is obtained as follows: $X_{ib}^* = X_i$ and Y_{ib}^* is generated with distribution function $F(\cdot \mid X_i, \widehat{\theta})$.

Stute (1997) suggests testing the specification of a regression model by comparing parametric and nonparametric estimations of the regression function. Specifically, if $m(\cdot, \theta_0)$ is the specified parametric regression function and

$$R_n(x) \equiv n^{-1}[\sum_{i=1}^{n} \{Y_i - m(X_i, \widehat{\theta})\} I(X_i \leq x)],$$

where $\widehat{\theta}$ is a root-n-consistent estimator of θ_0, then he proposes using the Cramér-von Mises-type statistic $T_n \equiv \sum_{l=1}^{n} R_n(X_l)^2$. This statistic cannot be directly applied to our problem since the specification that we consider is not a regression model if $J > 1$. However, observe that (2) holds if and only if

$$E(D_{ji} \mid X_i) = p_{ji}(\theta_0) \quad \text{for } j = 1, ..., J, \tag{10}$$

that is, our specification is equivalent to the the specification of J regression models. Hence, we can derive a test statistic for our problem following Stute's methodology as follows: first, we consider Stute's statistic for the j-th regression model in (10), which proves to be

$$C_{j,n}^{(ST)} \equiv n^{-2} \sum_{l=1}^{n} \left[\sum_{i=1}^{n} \{D_{ji} - p_{ji}(\widehat{\theta})\} I(X_i \leq X_l) \right]^2;$$

and then we consider the overall statistic

$$C_n^{(ST)} \equiv \sum_{j=1}^{J} C_{j,n}^{(ST)}.$$

This way of defining an overall statistic ensures that any deviation in any of the J regression models considered in (10) will be detected; but other definitions of the overall statistic would also ensure this, e.g. $\max_{1 \leq j \leq n} C_{j,n}^{(ST)}$ or $\sum_{j=1}^{J} \{C_{j,n}^{(ST)}\}^2$. The asymptotic null distribution of $C_n^{(ST)}$ is not known either, but approximate critical values can be derived using the bootstrap procedure described above. The asymptotic validity of this bootstrap procedure in this context can be proved using arguments similar to those in Stute et al. (1998).

Horowitz and Spokoiny (2001) propose testing the specification of a regression model comparing smoothed nonparametric and parametric estimations of the regression function with various bandwidths. Specifically, denote

$$R_{n,h}(x) \equiv \sum_{i=1}^{n} \{Y_i - m(X_i, \widehat{\theta})\} w_{i,h}(x),$$

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 the bandwidth and, as above, $m(\cdot, \theta_0)$ is the specified parametric regression function and $\widehat{\theta}$ is a root-n-consistent estimator of θ_0; with this notation, the statistic proposed in Horowitz and Spokoiny (2001) is $T_n \equiv \max_{h \in H_n}\{\sum_{l=1}^{n} R_{n,h}(X_l)^2 - \hat{N}_h\}/\hat{V}_h$, where $\hat{N}_h$ and $\hat{V}_h$ are normalizing constants and H_n is a finite set of possible bandwidth values. As above, we can apply this methodology in our context by first computing the statistic corresponding to the j-th regression model in (10), which proves to be

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

where $a_{il,h} \equiv \sum_{m=1}^{n} w_{i,h}(X_m) w_{l,h}(X_m)$ and $\hat{\sigma}_{ji}^2 \equiv p_{ji}(\widehat{\theta})\{1 - p_{ji}(\widehat{\theta})\}$; and then we consider the overall statistic

$$C_n^{(HS)} \equiv \sum_{j=1}^{J} C_{j,n}^{(HS)}.$$

Again, any deviation in any of the J regression models considered in (10) is detected in this way. The asymptotic null distribution of $C_n^{(HS)}$ is not known either, but approximate critical values can be derived using the bootstrap procedure described above. Observe that neither the bootstrap procedure nor the conditional variance estimators $\hat{\sigma}_{ji}^2$ that we use are the ones proposed in Horowitz and Spokoiny (2001), because we exploit the fact that in our case the dependent variable is binary; in this way, a better finite-sample performance is obtained.

4. SIMULATIONS

In order to check the behavior of the test statistics, we perform eight Monte Carlo experiments. In all of them we test the null hypothesis that (2) holds with the standard normal distribution as $F(\cdot)$. With these experiments we seek to examine whether the empirical size of the statistics is accurate, and also whether the statistics detect misspecification in the latent regression model due to non-linearities (Models 1-4), heteroskedasticity (Models 5-6) and non-normality in the distribution function $F(\cdot)$ (Models 7-8). In four experiments (Models 1, 3, 5, 7) the dependent variable Y only has two possible values (i.e. $J = 1$); in the other four Y has three possible values (i.e. $J = 2$). In six experiments (Models 1-2 and 5-8) only two regressors are included: a constant variable and a normal one; in the other two, an additional regressor is included in order to examine the extent to which results change if the number of regressors increases.

The specific models that we consider are as follows:

- *Model 1*: We generate n i.i.d. random vectors $\{(X_{2i}, u_i)'\}_{i=1}^n$, where X_{2i} and u_i are independent with standard normal distribution; then $Y_i^* = X_i'\beta_0 + c(X_{2i}^2 - 1) + u_i$, where $X_i = (1, X_{2i})'$, $\beta_0 = (0, 1)'$ and the value of c varies; finally $Y_i = 0$ if $Y_i^* < 0$, or 1 if $Y_i^* \geq 0$.

- *Model 2*: $\{(X_{2i}, u_i)'\}_{i=1}^n$ are generated as in Model 1; then $Y_i^* = X_i'\beta_0 + c(X_{2i}^2 - 1) + u_i$, where $X_i = (1, X_{2i})'$, $\beta_0 = (1, 1)'$ and the value of c varies; finally $Y_i = 0$ if $Y_i^* < 0$, or 1 if $0 \leq Y_i^* < \mu_0$, or 2 if $Y_i^* \geq \mu_0$, where $\mu_0 = 2$.

- *Model 3*: We generate n i.i.d. random vectors $\{(X_{2i}, X_{3i}, u_i)'\}_{i=1}^n$ all independent with standard normal distribution; then $Y_i^* = X_i'\beta_0 + c(X_{1i}^2 - 1)(X_{2i}^2 - 1) + u_i$, where $X_i = (1, X_{2i}, X_{3i})'$, $\beta_0 = (0, 1, 1)'$ and the value of c varies; finally $Y_i = 0$ if $Y_i^* < 0$, or 1 if $Y_i^* \geq 0$.

- *Model 4*: $\{(X_{2i}, X_{3i}, u_i)'\}_{i=1}^n$ are generated as in Model 3; then $Y_i^* = X_i'\beta_0 + c(X_{1i}^2 - 1)(X_{2i}^2 - 1) + u_i$, where $X_i = (1, X_{2i}, X_{3i})'$, $\beta_0 = (1, 1, 1)'$ and the value of c varies; finally $Y_i = 0$ if $Y_i^* < 0$, or 1 if $0 \leq Y_i^* < \mu_0$, or 2 if $Y_i^* \geq \mu_0$, where $\mu_0 = 2$.

- *Model 5*: Data are generated as in Model 1 with $c = 0$, but now the conditional distribution of u_i given $X_{2i} = x$ is normal with zero mean and variance $\exp(dX_{2i} - d^2/2)$ for various d; hence, the unconditional distribution of u_i has zero mean and unit variance.

- *Model 6*: Data are generated as in Model 2 with $c = 0$, but now u_i is generated as in Model 5.

- *Model 7*: Data are generated as in Model 1 with $c = 0$, but now $u_i = |d|^{1/2}\varepsilon_i - |d|^{-1/2}$ if $d > 0$, or $u_i = -|d|^{1/2}\varepsilon_i + |d|^{-1/2}$ if $d < 0$, and ε_i follows a gamma distribution with density function $f_\varepsilon(x) = x^{(1/|d|)-1}\exp(-x)/\Gamma(1/|d|)$ for various d; hence u_i has zero mean and unit variance. The limit distribution of u_i as d approaches 0 is the standard normal one, but if $|d|$ is large the distribution of u_i is highly skewed (for $d = 0$, we generate u_i from a standard normal distribution).

- *Model 8*: Data are generated as in Model 2 with $c = 0$, but now u_i is generated as in Model 7.

In all cases parameters are estimated by maximum likelihood assuming that the null hypothesis holds. Note that in Models 1-4 H_0 is true if and only if $c = 0$, whereas in Models 5-8 H_0 is true if and only if $d = 0$. In Tables 1-8 we report the proportions of rejections of H_0 at the 5% significance level for two different sample sizes: $n = 100$ and $n = 500$. When a bootstrap procedure is required we use $B = 100$ bootstrap replications. All results are based on 1000 simulation runs, performed using GAUSS programmes that are available from the authors on request. When computing $C_{n,l}^{(MP)}$ we consider various possible partitions of the support, but we only report the results for the statistic that yields the best performance, namely, $C_{n,3}^{(MP)}$ with $G = 2$ and $A_1 = \mathbb{R}\times(-\infty, 0)$ in Models 1-2 and 5-8; and $C_{n,3}^{(MP)}$ with $G = 4$ and $A_1 = \mathbb{R}\times(-\infty, 0) \times (-\infty, 0)$, $A_2 = \mathbb{R}\times(-\infty, 0) \times [0, \infty)$, $A_3 = \mathbb{R}\times[0, \infty) \times (-\infty, 0)$ in Models 3-4. When computing $C_n^{(HS)}$ we use a Gaussian kernel and, after some preliminary results, we decide to choose $H_n = \{\frac{i}{2}$, for $i = 1, ..., 5\}$ when $n = 100$ and $H_n = \{\frac{3i}{10}$, for $i = 1, ..., 5\}$ when $n = 500$.

First we discuss the results that we obtain for the statistics based on conditional moment restrictions. In all cases, the empirical size of the tests based on $C_{n,2}^{(M)}$ is much higher than the nominal level; this overrejection problem is especially severe when $n = 100$, and leads us to disregard $C_{n,2}^{(M)}$ as a test statistic. When we use $C_{n,1}^{(M)}$ the null hypothesis is also rejected too often; additionally, under the alternative the power of $C_{n,1}^{(M)}$ is much lower than that obtained with the other statistics. Therefore, the most reasonable moment-based statistic proves to be $C_{n,3}^{(M)}$, that is, it is crucial to take into account the specific nature of discrete choice models when estimating the covariance matrix $\mathbf{V}_0$. Also observe that increasing the degrees of freedom by partitioning the support of the independent variables does not usually produce an increase in the power of the tests; the only exception to this is when detecting non-normality (compare $C_{n,3}^{(M)}$ and $C_{n,3}^{(MP)}$ in Table 8). On the other hand, the test of overidentifying restrictions $C_n^{(BC)}$ yields worse results than the others, except again when detecting non-normal errors; in this case its performance is comparable to (or even better than) that of the others.

As regards the statistics based on the comparison between parametric and nonparametric estimations, all three behave reasonably well in terms of size. When comparing power performance, statistic $C_n^{(HS)}$ yields the best results in almost all cases. This is not a surprise in Models 1, 3, 5, 7, since they are all regression models, and Horowitz and Spokoiny (2001) derive the optimality of their statistic in this context; our results suggest that this optimality property also holds when it is applied in general ordered discrete choice models. The only exception to this rule is Model 8, where the generalization of Stute's statistic yields somewhat better results than $C_n^{(HS)}$. Comparing $C_n^{(AN)}$ and $C_n^{(ST)}$, the latter performs better when testing the specification of a binary choice models, but when $J \geq 2$ our results are not so conclusive: Andrew's statistic performs better than the generalization of Stute's statistic to detect non-linearities, but it performs worse with non-normal errors.

Finally, if we compare the two statistics that perform best from each approach, namely $C_{n,3}^{(M)}$ and $C_n^{(HS)}$, there is no clear-cut answer to the question of which of them performs better: the former performs better in detecting heteroskedasticity in the latent regression model, whereas the latter performs better in detecting non-normality or when the number of regressors is greater than one. Taking into account the huge difference in computation time between them, one might be tempted to say that $C_{n,3}^{(M)}$ should be preferred, but if computation time considerations are disregarded, then our results show that the gain in power that is obtained with the generalization of Horowitz-Spokoiny's statistic is enough reward for the additional programming effort.

5. EMPIRICAL APPLICATIONS

As an empirical illustration, we apply all the test statistics in two different contexts. First, we consider the well-known data on extramarital affairs described in Fair (1978). This is a famous data set that has become a milestone for models with qualitative dependent variable. The data come from a survey published in 1969 in *Psychology Today*, with questions about characteristics of the individual and the number of extramarital affairs (NEA) during the past year. The size of the sample is 601. As a dependent variable we consider: $Y = 0$ if NEA$=0$, $Y = 1$ if NEA$= 1, 2, 3$, or $Y = 2$ if NEA≥ 4. First we analyse whether an ordered probit specification is adequate for the conditional distribution of Y given as explanatory variables a constant variable, number of years married and sex (0 for female, 1 for male). Then we repeat the analysis using all possible explanatory variables, i.e. the three previous ones plus religiousness, education, whether there are children or not, age and self rating of marriage. In Table 9 we report the results of the ML estimation and in Table 10 we report the specification test statistics that are obtained. The statistic $C_{n,3}^{(MP)}$ is computed with $G = 2$ and partitioning the support of the regressors according to sex; on the other hand, the kernel weights required to obtain $C_n^{(HS)}$ are computed using a Gaussian kernel, regressors previously standardized to have unit variance, and the family of bandwidths $H_n = \{\frac{3i}{10}$, for $i = 1, ..., 5\}$. In all cases the bootstrap p-values are obtained with $B = 1000$ bootstrap replications. The results reported in Table 10 suggest that the ordered probit specification might not be adequate for $Y \mid X$ when X only includes a constant variable, number of years married and sex (the specification is rejected at the 10% level with most test statistics), but there is no evidence against it when X includes all possible explanatory variables.

As a second example, we consider the determinants of women's labour market participation and the type of participation (work full-time or part-time). Literature has broadly dealt with this topic and its influence on fertility and divorce rate, among other variables. We use data from the PSID (Wave 2001) to explain whether a woman does not work, works part-time or works full time. Our sample contains 2866 observations; this sample has been obtained considering only those women whose age is between 20 and 45, and whose marital status is other than "never married". Using information from the variable "Total hours of work (wife)", we define our dependent variable as follows: 0 if the woman reports 0 hours, 1 if she reports a positive number of hours, but less than a certain level ϕ, and 2 if the hours reported are greater than ϕ (specifically, we choose $\phi = 1440$ hours per year). We consider the following variables as determinants of this employment status: age (as a proxy for experience), age square (to reflect the non-linear influence of experience on employment status), whether the household contains children or not, education and the husband's labour income. In Table 10 we report the results of the ML estimation and the specification test statistics that are obtained. 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), and $C_n^{(HS)}$ is computed as in the previous application. Bootstrap p-values are again obtained with $B = 1000$ bootstrap replications. The results reported in Table 10 show that the ordered probit specification is rejected at the usual significance levels with almost all test statistics. However, if we consider a distribution function $F(\cdot)$ with fatter tails (specifically, a Student's t cdf with 5 degrees of freedom, re-scaled to have unit variance), results change dramatically: this specification is not rejected at the 5% level with most test statistics. Moreover, an adequate choice of $F(\cdot)$ is also crucial in analysing the influence of regressors on employment status: note that estimations with a Student's t cdf suggest that age, rather than husband's income, is the most significant explanatory variable.

6. CONCLUDING REMARKS

We discuss in this paper how to test the specification of ordered discrete choice models. Two main approaches can be followed: tests based on conditional moment restrictions and tests based on comparisons between parametric and nonparametric estimations. Our contribution in this paper is threefold: first, we propose a variant of the usual conditional moment statistics that exploits all the information in the model and is easy to compute (it is based on an artificial regression); second, we propose generalizations of the test statistics proposed in Stute (1997) and Horowitz and Spokoiny (2001), and describe how bootstrap critical values can be obtained for them in our context; and third, we compare the performance of these statistics (and some others) using various models that allow us to examine the empirical size of the tests and their ability to detect deviations from the assumptions that are usually made in applications: linearity, homoskedasticity and normality in the latent regression model. Our simulation results show that the behaviour of conditional moment tests crucially depends on how the covariance matrix is estimated; furthermore, if this estimator is adequately chosen, the resulting test statistic outperforms almost all other statistics considered here. On the other hand, our simulation results suggest that the optimality property of Horowitz-Spokoiny's statistic in regression models may also hold for the generalization of the statistic that we propose. Finally, the results of the empirical applications that we include to illustrate the performance of the statistics highlight the importance of selecting an accurate test statistic.

APPENDIX

A1. Analytical Expressions for the Estimators of $\mathbf{V}_0$ and $\mathbf{V}_0^{(P)}$

Hereafter, $f(\cdot)$ and $\dot{f}(\cdot)$ denote the first and second derivative of $F(\cdot)$, $f_{0i} \equiv f(-X_i'\beta)$, $\dot{f}_{0i} \equiv \dot{f}(-X_i'\beta)$ and, for $j = 1, ..., J-1$, $f_{ji} \equiv f(\mu_j - X_i'\beta)$, $\dot{f}_{ji} \equiv \dot{f}(\mu_j - X_i'\beta)$. Additionally, $f_{-1,i} \equiv 0$, $\dot{f}_{-1,i} \equiv 0$, $f_{Ji} \equiv 0$, $\dot{f}_{Ji} \equiv 0$ and $p_{ji} \equiv p_{ji}(\theta)$.

To compute $\mathbf{V}_{n,1}$ and $\mathbf{V}_{n,2}$, we must derive expressions for $\mathbf{B}_i(\theta)$, $\mathbf{g}_i(\theta)$ and $\mathbf{A}_i(\theta)$. Taking into account the definitions that are given 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 $-f_{ji}I(l=j) + f_{j-1,i}I(l=j-1)$. 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,

$$\mathbf{A}_i(\theta) \equiv \begin{pmatrix} \mathbf{A}_{11i}(\theta) & \mathbf{A}_{12i}(\theta) \\ \mathbf{A}_{12i}(\theta)' & \mathbf{A}_{22i}(\theta) \end{pmatrix},$$

where,

$$\mathbf{A}_{11i}(\theta) \equiv \left\{ \sum_{j=0}^{J} \frac{D_{ji}[(f_{ji} - f_{j-1,i})^2 - p_{ji}(\dot{f}_{ji} - \dot{f}_{j-1,i})]}{p_{ji}^2} \right\} X_i X_i',$$

$\mathbf{A}_{12i}(\theta)$ is the $k \times (J-1)$ matrix whose j-th column is

$$\left\{ \frac{D_{ji}[\dot{f}_{ji}p_{ji} - (f_{ji} - f_{j-1,i})f_{ji}]}{p_{ji}^2} + \frac{D_{j+1,i}[(f_{j+1,i} - f_{ji})f_{ji} - \dot{f}_{ji}p_{j+1,i}]}{p_{j+1,i}^2} \right\} X_i,$$

and $\mathbf{A}_{22i}(\theta)$ is the symmetric matrix whose $(j, j+l)$ element is

$$\left\{ \frac{D_{ji}(f_{ji}^2 - \dot{f}_{ji}p_{ji})}{p_{ji}^2} + \frac{D_{j+1,i}(f_{ji}^2 + \dot{f}_{ji}p_{j+1,i})}{p_{j+1,i}^2} \right\} I(l=0) - \frac{D_{j+1,i}f_{ji}f_{j+1,i}}{p_{j+1,i}^2} I(l=1).$$

To compute $\mathbf{V}_{n,3}$, we must derive expressions for $E_X\{\mathbf{m}_i(\theta)\mathbf{m}_i(\theta)'\}$, $E_X\{\mathbf{m}_i(\theta)\mathbf{g}_i(\theta)'\}$ and $E_X\{\mathbf{g}_i(\theta)\mathbf{g}_i(\theta)'\}$. In this case $E_X\{\mathbf{m}_i(\theta)\mathbf{m}_i(\theta)'\}$ 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 $-p_{ji}p_{li}$. On the other hand, $E_X\{\mathbf{m}_i(\theta)\mathbf{g}_i(\theta)'\} = -\mathbf{B}_i(\theta)$. Finally,

$$E_X\{\mathbf{g}_i(\theta)\mathbf{g}_i(\theta)'\} \equiv \begin{pmatrix} \mathbf{A}^*_{11i}(\theta) & \mathbf{A}^*_{12i}(\theta) \\ \mathbf{A}^*_{12i}(\theta)' & \mathbf{A}^*_{22i}(\theta) \end{pmatrix},$$

where $\mathbf{A}^*_{11i}(\theta) \equiv \{\sum_{j=0}^{J}(f_{ji}-f_{j-1,i})^2/p_{ji}\}X_iX_i'$, $\mathbf{A}^*_{12i}(\theta)$ 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 $\mathbf{A}^*_{22i}(\theta)$ is the symmetric matrix whose (j,l) element, for $l\geq j$, is $(1/p_{ji}+1/p_{j+1,i})f_{ji}^2 I(l=j)-(f_{ji}f_{j+1,i}/p_{j+1,i})I(l=j+1)$.

As for the estimators of $\mathbf{V}_0^{(P)}$, note that $\mathbf{V}_{n,1}^{(P)}$ and $\mathbf{V}_{n,2}^{(P)}$ can be computed using the above expressions. To compute $\mathbf{V}_{n,3}^{(P)}$, observe 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 $E_X\{\mathbf{m}_i^{(P)}(\theta)\mathbf{g}_i(\theta)'\} = -\mathbf{B}_i^{(P)}(\theta)$.

A2. Artificial Regressions to Compute $\mathbf{C}_{n,3}^{(M)}$ and $\mathbf{C}_{n,3}^{(MP)}$

Taking into account the expression for $\mathbf{V}_{n,3}$, in order to derive an artificial regression whose explained sum of squares is $C_{n,3}^{(M)}$, first we use a Cholesky decomposition of $n^{-1}\sum_{i=1}^{n}E_X\{\mathbf{m}_i(\widehat{\theta})\mathbf{m}_i(\widehat{\theta})'\}$ to derive the first set of explanatory variables in the artificial regression, and then we define the remaining explanatory variables and the dependent variable accordingly. Here we describe the artificial regression that is obtained in this way.

Denote $\widehat{p}_{ji}\equiv p_{ji}(\widehat{\theta})$ and $\widehat{\delta}_{ji}\equiv 1-F(\widehat{\mu}_j-X_i'\widehat{\beta})+F(-X_i'\widehat{\beta})$ 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:

$$\hat{\mathbf{c}}_{ji,l} \equiv \{\widehat{p}_{ji} I(l < j) + \widehat{\delta}_{j-1,i} I(l = j)\}/(\widehat{p}_{ji}\widehat{\delta}_{ji}\widehat{\delta}_{j-1,i})^{1/2},$$
$$\hat{\mathbf{d}}_{ji,l} \equiv \widehat{p}_{li}^{1/2}[-\widehat{p}_{ji} I(l < j) + \widehat{\delta}_{ji} I(l = j)]/(\widehat{\delta}_{li}\widehat{\delta}_{l-1,i})^{1/2},$$
$$\hat{\mathbf{e}}_{i,l} \equiv f(\widehat{\mu}_l - X_i'\widehat{\beta}) - f(\widehat{\mu}_{l-1} - X_i'\widehat{\beta}),$$
$$\hat{\mathbf{f}}_{ji,l} \equiv f(\widehat{\mu}_j - X_i'\widehat{\beta})\{I(l = j+1) - I(l = j)\}.$$

If we define $\mathbf{Z} \equiv [\mathbf{z}_1', ..., \mathbf{z}_n']'$, where $\mathbf{z}_i$ is the $J \times 1$ vector whose j-th element is $\hat{\mathbf{c}}_{ji}'\hat{\mathbf{m}}_i$, and $\mathbf{W}$ =$[\mathbf{W}^{(1)} : \mathbf{W}^{(2)} : \mathbf{W}^{(3)}]$, where $\mathbf{W}^{(l)} \equiv [\mathbf{w}_1^{(l)\prime}, ..., \mathbf{w}_n^{(l)\prime}]'$ 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 $\hat{\mathbf{c}}_{ji}'\hat{\mathbf{f}}_{li}$, then it follows that $\mathbf{Z}'\mathbf{W}^{(\mathbf{1})} = \sum_{i=1}^n \mathbf{m}_i(\widehat{\theta})'$, $\mathbf{Z}'\mathbf{W}_* = -\sum_{i=1}^n \mathbf{g}_i(\widehat{\theta}) = \mathbf{0}$, where $\mathbf{W}_* \equiv [\mathbf{W}^{(\mathbf{2})} : \mathbf{W}^{(\mathbf{3})}]$, and $\mathbf{W}^{(\mathbf{1})\prime}\mathbf{W}^{(\mathbf{1})} - \mathbf{W}^{(\mathbf{1})\prime}\mathbf{W}_*(\mathbf{W}_*'\mathbf{W}_*)^{-1}\mathbf{W}_*'\mathbf{W}^{(\mathbf{1})} = n\mathbf{V}_{n,3}$. Therefore $C_{n,3}^{(M)}$ coincides with the explained sum of squares in the artificial regression with vector of dependent observations $\mathbf{Z}$ and matrix of observations $\mathbf{W}$. Similarly, $C_{n,3}^{(MP)}$ coincides with the explained sum of squares in the artificial regression with vector of dependent observations $\mathbf{Z}^{(P)} \equiv [\mathbf{z}_1^{(P)\prime}, ..., \mathbf{z}_n^{(P)\prime}]'$ and matrix of observations $\mathbf{W}^{(P)} \equiv [\mathbf{W}^{(P1)} : \mathbf{W}^{(P2)} : \mathbf{W}^{(P3)}]$, where $\mathbf{z}_i^{(P)} \equiv \mathbf{z}_i \otimes \mathbf{P}_i$, $\mathbf{W}^{(Pl)}$=$[\mathbf{w}_1^{(Pl)\prime}, ..., \mathbf{w}_n^{(Pl)\prime}]'$ 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 $\mathbf{w}_i^{(P3)} \equiv \mathbf{w}_i^{(3)} \otimes \mathbf{P}_i$.

REFERENCES

Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Cambridge.

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.

Calhoun, C. A. (1989). "Estimating the Distribution of Desired Family Size and Excess Fertility", *Journal of Human Resources*, Vol. 24, pp. 709-724.

Glewwe, P. (1997). "A Test of the Normality Assumption in the Ordered Probit Model". *Econometric Reviews*, Vol. 16, pp. 1-19

Gustaffson, S. and Stafford, F. (1992). "Child Care Subsidies and Labor Supply in Sweden", *Journal of Human Resources*, Vol. 27, pp. 204-230.

Hamilton, J. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

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.

Jiménez, E. and Kugler, B. (1987). "The Earnings Impact of Training Duration in a Developing Country: An Ordered Probit Selection Model of Colombia's Servicio Nacional de Aprendizaje", *Journal of Human Resources*, Vol. 22, pp. 228-247.

Klein, R.W. and Sherman, R.P. (1997). "Estimating New Product Demand from Biased Survey Data", *Journal of Econometrics*, Vol. 76, pp. 53-76.

Klein, R.W. and Sherman, R.P. (2002). "Shift Restrictions and Semiparametric Estimation in Ordered Response Models", *Econometrica*, Vol. 70, pp. 663-691.

MacKinnon, J. G. (1992). "Model Specification Tests and Artificial Regressions", *Journal of Economic Literature*, Vol. 30, pp. 102-146.

Matzkin, R. (1992). "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold-Crossing Models", *Econometrica*, Vol. 60, pp. 239-270.

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.

Santos Silva, J. M. C. (2001). "A Score Test for Non-Nested Hypotheses with Applications to Discrete Data Models", *Journal of Applied Econometrics*, Vol. 16, pp. 577-597.

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.

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.

TABLE 1

Proportion of Rejections of $\mathbf{H}_0$ in Model 1, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_n^{(AN)}$	$C_n^{(ST)}$	$C_n^{(HS)}$
c				$n = 100$			
-0.8	0.728	0.999	0.995	0.989	0.935	0.978	0.992
-0.6	0.852	0.970	0.940	0.928	0.764	0.880	0.920
-0.4	0.687	0.757	0.682	0.614	0.358	0.541	0.650
-0.2	0.275	0.323	0.221	0.188	0.107	0.166	0.209
0.0	0.100	0.123	0.047	0.056	0.052	0.056	0.057
0.2	0.267	0.316	0.213	0.168	0.185	0.196	0.188
0.4	0.712	0.764	0.687	0.639	0.562	0.599	0.621
0.6	0.862	0.964	0.946	0.912	0.870	0.907	0.917
0.8	0.729	0.988	0.984	0.985	0.973	0.984	0.988
c				$n = 500$			
-0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.4	0.999	0.999	0.999	0.998	0.961	0.993	0.995
-0.2	0.707	0.736	0.722	0.640	0.299	0.538	0.544
0.0	0.065	0.070	0.051	0.048	0.043	0.054	0.055
0.2	0.747	0.772	0.758	0.648	0.474	0.575	0.525
0.4	0.998	0.998	0.998	0.998	0.991	0.996	0.992
0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 2

Proportion of Rejections of $\mathbf{H}_0$ in Model 2, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_n^{(BC)}$	$C_n^{(AN)}$	$C_n^{(ST)}$	$C_n^{(HS)}$
c				$n = 100$				
-0.8	0.290	0.994	0.987	0.996	0.580	0.970	0.741	0.980
-0.6	0.321	0.935	0.933	0.939	0.637	0.907	0.671	0.917
-0.4	0.253	0.735	0.721	0.704	0.555	0.663	0.542	0.721
-0.2	0.118	0.330	0.272	0.253	0.217	0.238	0.226	0.293
0.0	0.067	0.120	0.048	0.044	0.072	0.069	0.049	0.060
0.2	0.097	0.302	0.252	0.233	0.340	0.207	0.189	0.240
0.4	0.242	0.737	0.729	0.725	0.696	0.621	0.440	0.702
0.6	0.297	0.935	0.935	0.950	0.754	0.896	0.716	0.918
0.8	0.252	0.990	0.986	0.995	0.674	0.983	0.871	0.986
c				$n = 500$				
-0.8	0.999	1.000	1.000	1.000	0.999	1.000	1.000	1.000
-0.6	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.4	0.976	0.996	0.997	1.000	1.000	1.000	0.997	1.000
-0.2	0.703	0.860	0.858	0.874	0.831	0.784	0.729	0.843
0.0	0.064	0.079	0.050	0.053	0.052	0.045	0.060	0.052
0.2	0.718	0.856	0.875	0.893	0.859	0.745	0.621	0.811
0.4	0.989	0.999	1.000	1.000	0.999	0.999	0.998	1.000
0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000

TABLE 3

Proportion of Rejections of $\mathbf{H}_0$ in Model 3, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_n^{(AN)}$	$C_n^{(ST)}$	$C_n^{(HS)}$
c				$n = 100$			
-0.8	0.533	0.606	0.561	0.561	0.467	0.640	0.852
-0.6	0.480	0.558	0.497	0.468	0.330	0.428	0.656
-0.4	0.324	0.441	0.307	0.272	0.188	0.217	0.358
-0.2	0.185	0.261	0.126	0.093	0.085	0.079	0.144
0.0	0.109	0.170	0.052	0.054	0.057	0.061	0.066
0.2	0.170	0.247	0.105	0.099	0.087	0.103	0.168
0.4	0.338	0.453	0.322	0.285	0.107	0.154	0.402
0.6	0.476	0.550	0.469	0.454	0.179	0.276	0.661
0.8	0.500	0.569	0.536	0.539	0.248	0.402	0.856
c				$n = 500$			
-0.8	0.505	0.615	0.832	0.981	0.995	1.000	1.000
-0.6	0.430	0.519	0.752	0.941	0.958	0.992	1.000
-0.4	0.476	0.496	0.617	0.743	0.693	0.800	0.960
-0.2	0.324	0.366	0.340	0.263	0.197	0.206	0.394
0.0	0.062	0.072	0.050	0.048	0.067	0.059	0.072
0.2	0.298	0.329	0.291	0.239	0.116	0.156	0.390
0.4	0.457	0.475	0.598	0.720	0.394	0.560	0.953
0.6	0.426	0.482	0.726	0.927	0.751	0.948	1.000
0.8	0.498	0.590	0.815	0.978	0.950	0.998	1.000

TABLE 4

Proportion of Rejections of $\mathbf{H}_0$ in Model 4, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_n^{(BC)}$	$C_n^{(AN)}$	$C_n^{(ST)}$	$C_n^{(HS)}$
c				$n = 100$				
-0.8	0.280	0.526	0.393	0.373	0.350	0.510	0.334	0.739
-0.6	0.251	0.505	0.388	0.303	0.400	0.374	0.257	0.551
-0.4	0.216	0.437	0.311	0.248	0.350	0.218	0.157	0.302
-0.2	0.137	0.284	0.154	0.123	0.215	0.110	0.092	0.107
0.0	0.109	0.214	0.054	0.053	0.102	0.050	0.065	0.047
0.2	0.142	0.288	0.156	0.143	0.113	0.082	0.104	0.212
0.4	0.215	0.434	0.302	0.231	0.171	0.147	0.185	0.496
0.6	0.252	0.522	0.387	0.318	0.234	0.180	0.251	0.670
0.8	0.269	0.531	0.428	0.401	0.204	0.241	0.246	0.766
c				$n = 500$				
-0.8	0.606	0.812	0.862	0.962	0.906	1.000	0.999	1.000
-0.6	0.569	0.699	0.819	0.883	0.920	0.997	0.981	0.999
-0.4	0.520	0.601	0.732	0.703	0.858	0.847	0.764	0.973
-0.2	0.319	0.422	0.437	0.302	0.451	0.299	0.252	0.384
0.0	0.081	0.098	0.047	0.055	0.050	0.062	0.051	0.060
0.2	0.308	0.407	0.422	0.286	0.283	0.196	0.230	0.676
0.4	0.543	0.622	0.743	0.697	0.734	0.582	0.630	0.995
0.6	0.547	0.675	0.820	0.864	0.855	0.835	0.803	1.000
0.8	0.585	0.763	0.844	0.951	0.830	0.957	0.871	1.000

TABLE 5

Proportion of Rejections of $\mathbf{H}_0$ in Model 5, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_n^{(AN)}$	$C_n^{(ST)}$	$C_n^{(HS)}$
d				$n = 100$			
-0.8	0.564	0.613	0.479	0.383	0.356	0.395	0.406
-0.6	0.382	0.438	0.318	0.251	0.231	0.243	0.252
-0.4	0.223	0.266	0.165	0.135	0.138	0.147	0.155
-0.2	0.143	0.169	0.083	0.075	0.080	0.074	0.074
0.0	0.100	0.123	0.047	0.056	0.052	0.056	0.057
0.2	0.124	0.163	0.065	0.057	0.066	0.074	0.083
0.4	0.240	0.285	0.181	0.130	0.087	0.145	0.158
0.6	0.397	0.443	0.315	0.253	0.130	0.225	0.275
0.8	0.535	0.597	0.460	0.371	0.204	0.355	0.424
d				$n = 500$			
-0.8	0.988	0.990	0.988	0.975	0.918	0.957	0.942
-0.6	0.930	0.936	0.924	0.868	0.711	0.815	0.756
-0.4	0.663	0.686	0.657	0.539	0.371	0.486	0.413
-0.2	0.251	0.266	0.227	0.182	0.127	0.151	0.133
0.0	0.065	0.070	0.051	0.048	0.043	0.054	0.055
0.2	0.231	0.248	0.217	0.154	0.086	0.156	0.149
0.4	0.649	0.666	0.652	0.530	0.238	0.455	0.438
0.6	0.917	0.926	0.913	0.865	0.508	0.768	0.763
0.8	0.988	0.993	0.988	0.980	0.775	0.945	0.954

TABLE 6

Proportion of Rejections of $\mathbf{H}_0$ in Model 6, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_{n}^{(BC)}$	$C_{n}^{(AN)}$	$C_{n}^{(ST)}$	$C_{n}^{(HS)}$
d				$n = 100$				
-0.8	0.442	0.588	0.429	0.437	0.056	0.227	0.187	0.389
-0.6	0.270	0.405	0.272	0.283	0.049	0.151	0.123	0.230
-0.4	0.142	0.263	0.152	0.153	0.059	0.105	0.079	0.132
-0.2	0.073	0.155	0.073	0.072	0.063	0.070	0.070	0.077
0.0	0.067	0.120	0.048	0.044	0.072	0.069	0.049	0.060
0.2	0.082	0.160	0.065	0.060	0.062	0.069	0.066	0.073
0.4	0.156	0.263	0.155	0.132	0.076	0.066	0.104	0.122
0.6	0.259	0.408	0.265	0.265	0.067	0.091	0.147	0.192
0.8	0.430	0.603	0.443	0.441	0.065	0.136	0.211	0.283
d				$n = 500$				
-0.8	0.995	0.996	0.996	0.989	0.061	0.819	0.872	0.974
-0.6	0.954	0.967	0.943	0.931	0.058	0.488	0.558	0.832
-0.4	0.665	0.712	0.652	0.610	0.057	0.191	0.228	0.461
-0.2	0.193	0.217	0.183	0.181	0.050	0.087	0.087	0.130
0.0	0.064	0.079	0.050	0.053	0.052	0.045	0.060	0.052
0.2	0.216	0.233	0.207	0.161	0.053	0.071	0.092	0.138
0.4	0.650	0.696	0.634	0.597	0.047	0.134	0.208	0.378
0.6	0.963	0.973	0.963	0.938	0.049	0.327	0.513	0.703
0.8	0.995	0.998	0.998	0.997	0.054	0.648	0.798	0.928

TABLE 7

Proportion of Rejections of $\mathbf{H}_0$ in Model 7, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_n^{(AN)}$	$C_n^{(ST)}$	$C_n^{(HS)}$
d				$n = 100$			
-0.8	0.486	0.550	0.308	0.254	0.142	0.287	0.342
-0.6	0.400	0.450	0.244	0.187	0.103	0.207	0.232
-0.4	0.273	0.337	0.185	0.118	0.086	0.164	0.181
-0.2	0.182	0.218	0.097	0.090	0.064	0.104	0.114
0.0	0.100	0.123	0.047	0.056	0.052	0.056	0.057
0.2	0.175	0.209	0.085	0.088	0.094	0.094	0.100
0.4	0.287	0.335	0.166	0.125	0.148	0.166	0.161
0.6	0.397	0.469	0.210	0.184	0.231	0.246	0.263
0.8	0.503	0.564	0.324	0.217	0.273	0.299	0.310
d				$n = 500$			
-0.8	0.985	0.991	0.970	0.950	0.682	0.943	0.930
-0.6	0.948	0.953	0.899	0.834	0.508	0.820	0.796
-0.4	0.831	0.850	0.768	0.658	0.345	0.605	0.555
-0.2	0.495	0.513	0.420	0.347	0.163	0.333	0.294
0.0	0.065	0.070	0.051	0.048	0.043	0.054	0.055
0.2	0.488	0.515	0.418	0.335	0.251	0.347	0.303
0.4	0.799	0.820	0.726	0.648	0.495	0.616	0.567
0.6	0.961	0.968	0.925	0.837	0.732	0.836	0.799
0.8	0.985	0.990	0.966	0.947	0.880	0.939	0.928

TABLE 8

Proportion of Rejections of $\mathbf{H}_0$ in Model 8, $\alpha = 0.05$

	$C_{n,1}^{(M)}$	$C_{n,2}^{(M)}$	$C_{n,3}^{(M)}$	$C_{n,3}^{(MP)}$	$C_n^{(BC)}$	$C_n^{(AN)}$	$C_n^{(ST)}$	$C_n^{(HS)}$
d				$n = 100$				
-0.8	0.350	0.601	0.393	0.492	0.515	0.388	0.668	0.589
-0.6	0.256	0.444	0.274	0.395	0.488	0.307	0.534	0.478
-0.4	0.198	0.354	0.233	0.256	0.359	0.196	0.412	0.316
-0.2	0.096	0.218	0.104	0.159	0.218	0.129	0.266	0.208
0.0	0.067	0.120	0.048	0.044	0.062	0.069	0.049	0.060
0.2	0.109	0.210	0.111	0.137	0.120	0.147	0.230	0.200
0.4	0.172	0.332	0.175	0.274	0.251	0.266	0.393	0.351
0.6	0.258	0.483	0.280	0.408	0.331	0.361	0.546	0.494
0.8	0.346	0.597	0.390	0.505	0.369	0.459	0.646	0.576
d				$n = 500$				
-0.8	0.899	0.974	0.964	1.000	0.984	0.993	1.000	0.999
-0.6	0.777	0.886	0.870	0.994	0.960	0.962	0.998	0.996
-0.4	0.589	0.739	0.710	0.945	0.900	0.808	0.946	0.937
-0.2	0.292	0.374	0.313	0.655	0.638	0.483	0.815	0.617
0.0	0.064	0.079	0.050	0.053	0.052	0.045	0.060	0.052
0.2	0.321	0.399	0.358	0.634	0.572	0.489	0.754	0.580
0.4	0.592	0.722	0.686	0.944	0.826	0.869	0.978	0.920
0.6	0.786	0.914	0.891	0.997	0.936	0.971	0.998	0.989
0.8	0.903	0.969	0.958	1.000	0.980	0.999	1.000	0.998

TABLE 9: Extramarital Affairs Data

Ordered Probit Model for $Y \mid X$

	$X = (X_1, X_2, X_3)$	$X = (X_1, ..., X_8)$
	Estimates (with t-statistics)	
Constant	-1.060 (-9.077)	0.720 (1.466)
Years of Marr.	0.040 (3.976)	0.062 (3.397)
Sex	0.096 (0.883)	0.133 (1.068)
Religiousness		-0.202 (-4.045)
Education		0.021 (0.813)
Child./No Child.		0.149 (0.942)
Age		-0.235 (-2.342)
S.R.M.		-0.276 (-5.426)
	P-values of test statistics	
$C_{n,1}^{(M)}$	0.307	0.421
$C_{n,2}^{(M)}$	0.063	0.372
$C_{n,3}^{(M)}$	0.076	0.365
$C_{n,3}^{(MP)}$	0.074	0.361
$C_n^{(BC)}$	0.082	0.134
$C_n^{(AN)}$	0.188	0.109
$C_n^{(ST)}$	0.115	0.348
$C_n^{(HS)}$	0.154	0.435

TABLE 10: Female Labour Participation Data

Ordered Discrete Choice Model for $Y \mid X$

	Normal cdf	Student's t_5 cdf
	Estimates (with t-statistics)	
Constant	-0.553 (-0.839)	-0.751 (-1.163)
Age	0.054 (1.253)	0.067 (1.715)
Age^2	-0.004 (-0.586)	-0.008 (-11.642)
Child./No Child.	-0.226 (-8.176)	-0.181 (-8.780)
Education	0.074 (6.515)	0.059 (6.707)
Husband's Income	-0.009 (-10.103)	-0.001 (-1.140)
	P-values of test statistics	
$C_{n,1}^{(M)}$	0.002	0.227
$C_{n,2}^{(M)}$	0.001	0.215
$C_{n,3}^{(M)}$	0.001	0.173
$C_{n,3}^{(MP)}$	0.003	0.268
$C_{n}^{(BC)}$	0.002	0.010
$C_{n}^{(AN)}$	0.002	0.009
$C_{n}^{(ST)}$	0.125	0.121
$C_{n}^{(HS)}$	0.013	0.148