Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Specification tests in ordered logit and probit models

Andrew A. Weiss

ECONOMETRIC REVIEWS, 16(4), 361-391 (1997)

SPECIFICATION TESTS IN ORDERED LOGIT AND PROBIT MODELS

Andrew A. Weiss

Econometrics Group
Faculty of Commerce and Administration
Victoria University of Wellington
P.O. Box 600, Wellington
New Zealand

Key Words and Phrases: Lagrange multiplier; Information matrix; Chi-squared.

JEL Classification: C12, C25, C52

ABSTRACT

In this paper, I study the application of various specification tests to ordered logit and probit models with heteroskedastic errors, with the primary focus on the ordered probit model. The tests are Lagrange multiplier tests, information matrix tests, and chi-squared goodness of fit tests. The alternatives are omitted variables in the regression equation, omitted variables in the equation describing the heteroskedasticity, and non-logistic/non-normal errors. The alternative error distributions include a generalized logistic distribution in the ordered logit model and the Pearson family in the ordered

probit model. I also consider the behavior of the tests under sequences of local alternatives and use Monte Carlo methods to study the small sample properties of the tests in particular models. In the LM tests, I examine how closely the asymptotic distributions under the local alternatives approximate the small sample distributions of the test statistics.

1. Introduction.

Qualitative response models are models in which the observed dependent variable takes on discrete values. These models can often be thought of as arising from the incomplete observation of an underlying continuous variable. Ordered logit (OL) and ordered probit (OP) models are obtained when the distributions of underlying continuous variables are logistic and normal, respectively, and larger values of the dependent variables correspond to the underlying continuous variables lying further to the right in partitions of the real line.[1]

Qualitative response models are typically estimated by maximum likelihood (ML) and, as is well known, misspecification of the model generally leads to inconsistent estimators of the parameters. Because of this, considerable attention has been paid to constructing tests for the specification of qualitative response models (e.g., Bera, Jarque and Lee 1984, Davidson and MacKinnon 1984, Chesher and Irish 1987, and Pagan and Vella 1989). The emphasis has been on tests that are easy to perform, such as Lagrange multiplier (LM) and conditional moment tests.

In this paper, I study the application of these testing principles to OL and OP models with heteroskedastic errors. I consider tests for omitted

[1] Greene (1993, p 672) lists some empirical examples of ordered models that have appeared in the literature.

variables in the regression equation and in the equation describing the heteroskedasticity, and for non-logistic/non-normal errors. The alternative error distributions include a generalized logistic distribution in the OL model and the Pearson family in the OP model. I derive the forms of the LM tests, and then show how these relate to the information matrix (IM) test (White 1982). I also consider the behavior of the tests under sequences of local alternatives and study the small sample properties of the tests in particular models using Monte Carlo methods. In the LM tests, I also examine how closely the asymptotic distributions under the local alternatives approximate the small sample distributions of the test statistics.

In section 2 of the paper I define the basic ordered qualitative response model and recall the properties of the ML estimator of its parameters. The general forms of the LM and IM tests are given in section 3; while sections 4 and 5 contain more detailed discussions of the tests in the OP and OL models, respectively. The behavior under local alternatives is summarized in section 6 and the results of the Monte Carlo experiment are presented in section 7. Concluding comments are given in section 8.

2. The Model.

The observed dependent variable is denoted by y_i, $i = 1, \ldots, n$, where n is the number of observations. Its values are determined by an unobserved continuous variable, y_i^*, and a partition of the real line. Let

$$y_i = j \quad \text{if and only if} \quad \gamma_j^* \leq y_i^* < \gamma_{j+1}^* \quad j = 0, \ldots, k, \tag{1}$$

where the γ_j^* are thresholds determining the partition. The regression equation for y_i^* is given by

$$y_i^* = x_i'\beta^* + \epsilon_i, \tag{2}$$

where x_i is a vector of explanatory variables and β^* is a vector of parameters. The error term ϵ_i has variance $\sigma_0^2\sigma_i^{*2}$, where

$$\sigma_i^{*2} = h(\alpha_0^* + z_i'\alpha^*), \tag{3}$$

h is a given function, z_i is a vector of explanatory variables, α_0^* is normalized such that $h(\alpha_0^*) = 1$, and σ_0^2 is determined by F, the c.d.f. of ϵ_i/σ_i^*.[2] Leading examples of the model are the OP and OL models. In the former, $\epsilon_i \sim N(0, \sigma_i^{*2})$ and $\sigma_0^2 = 1$; while in the latter, ϵ_i/σ_i^* is distributed as logistic, $\sigma_0^2 = \pi^2/3$, and

$$F(u) = \Lambda(u) \equiv (1 + exp\{-u\})^{-1}.$$

Common examples for h are $h(u) = u^2$ and $h(u) = exp(2u)$. γ_0^* is set to $-\infty$, γ_{k+1}^* is set to $+\infty$, and, assuming that β^* includes an intercept, γ_1^* is normalized to zero. The variables in x_i are linearly independent, as are those in z_i, and the observations are independent of one another.

From the distributional assumption,

$$P(y_i = j) = F\left(\frac{\gamma_{j+1}^* - x_i'\beta^*}{\sigma_i^*}\right) - F\left(\frac{\gamma_j^* - x_i'\beta^*}{\sigma_i^*}\right).$$

Hence, the log-likelihood function is given by

$$L(\theta_1) = \sum_{i=1}^{n}\sum_{j=0}^{k} 1(y_i = j)\ \log\ \left[F\left(\frac{\gamma_{j+1} - x_i'\beta}{\sigma_i}\right) - F\left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right)\right], \tag{4}$$

[2] F is assumed to be twice continuously differentiable, and $0 < F(u) < 1$ and $\partial F(u)/\partial u > 0$ for every u. $h(u)$ is twice continuously differentiable in u and $h(u) > 0$ for all u. More generally, the model is assumed to satisfy conditions like Assumptions 9.2.1-9.2.3 of Amemiya (1985, p 270).

where $1(u)$ is the indicator function of the event u (equal to one if u occurs and zero otherwise), $\sigma_i = h(\alpha_0^* + z_i'\alpha)^{\frac{1}{2}}$, and θ_1 is the vector of parameters. That is, $\theta_1 = (\beta', \alpha', \gamma')'$, where $\gamma = (\gamma_2, \ldots, \gamma_k)'$. The vector of true parameters is denoted by θ_1^*.

Let $1_i^{(j)} = 1(y_i = j)$, $F_i^{(j)} = F\big((\gamma_j - x_i'\beta)/\sigma_i\big)$, and $P_i^{(j)} = F_i^{(j+1)} - F_i^{(j)}$. Then following Amemiya (1985), the maximum likelihood estimate (MLE), denoted by $\hat{\theta}_1$, is a solution of

$$\frac{\partial L}{\partial \theta_1} = \sum_{i=1}^{n} \sum_{j=0}^{k} \frac{1_i^{(j)}}{P_i^{(j)}} \frac{\partial P_i^{(j)}}{\partial \theta_1} = 0. \tag{5}$$

As discussed in Amemiya (1985, p. 271), a solution to (5) may not exist (see also Albert and Anderson 1984). For example, if x_i includes a dummy variable that is equal to one if and only if $1_i^{(k)} = 1$, then maximizing $L(\theta_1)$ will send the coefficient on the dummy variable to $+\infty$. Such cases are ruled out, implying that $\hat{\theta}_1$ exists for n large enough. When the log-likelihood function is globally concave (e.g., Pratt 1981), $\hat{\theta}_1$ also maximizes $L(\theta_1)$.

Now, $E[1_i^{(j)} | x_i, z_i] = P(y_i = j)$ and, evaluated at θ_1^*, $P_i^{(j)} = P(y_i = j)$. Hence,

$$E\left[\frac{\partial L(\theta_1)}{\partial \theta_1}\right]\bigg|_{\theta_1^*} = \sum_{i=1}^{n} \sum_{j=0}^{k} E\left[\frac{\partial P_i^{(j)}}{\partial \theta_1}\right]\bigg|_{\theta_1^*} = 0, \tag{6}$$

where the last equality follows because the probabilities must sum to 1. The information matrix, denoted by $\mathbf{I}_{11}$, is obtained by writing the score vector in (5) as

$$\frac{\partial L}{\partial \theta_1} = \sum_{i=1}^{n} \bar{P}_i \bar{1}_i,$$

where

$$\bar{P}_i = \left(\frac{1}{P_i^{(0)}} \frac{\partial P_i^{(0)}}{\partial \theta_1}, \ldots, \frac{1}{P_i^{(k)}} \frac{\partial P_i^{(k)}}{\partial \theta_1}\right) \tag{7}$$

and $\bar{1}_i = \left(1_i^{(0)}, \ldots, 1_i^{(k)}\right)'$. Then since

$$\mathbf{I}_{11} = E\left[-n^{-1}\frac{\partial^2 L(\theta_1)}{\partial\theta_1\partial\theta_1'}\right]\Bigg|_{\theta_1^*} = E\left[n^{-1}\frac{\partial L(\theta_1)}{\partial\theta_1}\frac{\partial L(\theta_1)}{\partial\theta_1'}\right]\Bigg|_{\theta_1^*}$$

when the model is correctly specified, it follows that

$$\mathbf{I}_{11} = n^{-1}\sum_{i=1}^{n} E\left[\bar{P}_i\bar{1}_i\bar{1}_i'\bar{P}_i'\right]\big|_{\theta_1^*} = n^{-1}\sum_{i=1}^{n} E\left[\bar{P}_i I_i \bar{P}_i'\right]\big|_{\theta_1^*}, \tag{8}$$

where $I_i = E\left[\bar{1}_i\bar{1}_i' \,\middle|\, x_i, z_i\right]$ is the diagonal matrix with diagonal elements $P(y_i = j)$. Assuming that $\mathbf{I}_{11}$ is positive definite, equation (6) implies that θ_1^* corresponds to a local maximum of $E[L(\theta_1)]$. This implies that one of the roots of (5) will be consistent and, if $\hat{\theta}_1$ is the consistent root, that

$$n^{\frac{1}{2}}(\hat{\theta}_1 - \theta_1^*) \xrightarrow{d} N(0, \mathbf{I}_{11}^{-1}).$$

3. The Lagrange Multiplier and Information Matrix Tests – General Forms.

The LM test is based on testing a set of restriction in a more general form of the model. To augment the model, let

$$y_i^* = x_i'\beta^* + w_i'b^* + \epsilon_i \tag{9}$$

and

$$\sigma_i^{*2} = h(\alpha_0^* + z_i'\alpha^* + \nu_i'a^*), \tag{10}$$

where w_i and ν_i contain extra explanatory variables. The restrictions on (9) and (10) are that $b^* = 0$ and $a^* = 0$.[3] As noted in the Introduction, the more

[3] Of course, non-linear models and restrictions could be considered, but linear models and exclusion restrictions are the most common and the basic forms of the tests would be the same as those given here.

general form of the distribution of ϵ_i/σ_i^* depends on the model. These alternatives are detailed in section 4 and 5 below. Generically, the distribution is parameterized by a vector c and written as F_c, and the restrictions are given by $c = 0$. The full parameter vector is $\theta = (\theta_1', \theta_2)'$, where $\theta_2 = (b', a', c')'$; and when all the restrictions are imposed, the vector of parameters to be estimated is θ_1.

The log-likelihood function is given by

$$L_c(\theta) = \sum_{i=1}^{n} \sum_{j=0}^{k} 1_i^{(j)} \log(F_{ci}^{(j+1)} - F_{ci}^{(j)}), \tag{11}$$

where $F_{ci}^{(j)} = F_c\left((\gamma_j - x_i'\beta - w_i'b)/\sigma_i\right)$ and $\sigma_i = h(\alpha_0^* + z_i'\alpha + \nu_i'a)^{\frac{1}{2}}$; and under standard regularity conditions (e.g., Bera et al. 1984), the LM test statistic is given by

$$\xi_{LM} = n^{-1} \widehat{\frac{\partial L_c}{\partial \theta'}} \hat{\Sigma}^{-1} \widehat{\frac{\partial L_c}{\partial \theta}} \overset{d}{\dashrightarrow} \chi^2_{\ell_2}, \tag{12}$$

where the "hat" means the term is evaluated at the restricted MLE $(\hat{\theta}_1', 0)'$, ℓ_2 is the number of parameters in θ_2, and

$$\Sigma = E\left[n^{-1} \frac{\partial L_c(\theta)}{\partial \theta} \frac{\partial L_c(\theta)}{\partial \theta'}\right]\Bigg|_{(\theta_1^{*\prime}, 0)'}.$$

The expression for $\partial L_c/\partial \theta$ is equivalent to (5) with $P_i^{(j)}$ defined using (9), (10), and F_c rather than (2), (3), and F. Similarly, expressions for Σ are equivalent to the second and third terms in (8). Estimates of Σ are obtained from (8) by replacing the expectations by the corresponding sample terms evaluated at the restricted MLE. The second term in (8) corresponds to using the outer product of gradient (OPG) form of Σ and the third to using the scoring form. As numerous authors (e.g., Davidson and MacKinnon 1983) have noted, the use of OPG estimators in LM tests frequently lead to tests with poor small sample properties. This also occurred in the Monte Carlo experiment performed for this paper.

Consider now the IM test. From White (1982), this is based on an estimate of the difference

$$\text{vech}\left\{ E\left[n^{-1}\frac{\partial L(\theta_1)}{\partial \theta_1}\frac{\partial L(\theta_1)}{\partial \theta_1'}\right]\bigg|_{\theta_1^*} - E\left[-n^{-1}\frac{\partial^2 L(\theta_1)}{\partial \theta_1 \partial \theta_1'}\right]\bigg|_{\theta_1^*}\right\}.$$

Note that the log-likelihood function is L and the parameter vector is θ_1 – the model corresponds to the null hypothesis in the LM test. This enables the direct comparison of the IM and LM tests. The difference is estimated by

$$n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{k}\frac{1_i^{(j)}}{P_i^{(j)}}\text{vech}\frac{\partial^2 P_i^{(j)}}{\partial \theta_1 \partial \theta_1'}, \tag{13}$$

and taking expectations in (13) gives

$$n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{k}E\left[\frac{P(y_i = j)}{P_i^{(j)}}\text{vech}\frac{\partial^2 P_i^{(j)}}{\partial \theta_1 \partial \theta_1'}\right]\bigg|_{\theta_1^*}. \tag{14}$$

Under the null that the model is correctly specified and at $\theta_1 = \theta_1^*$, $P_i^{(j)} = P(y_i = j)$ and (14) is equal to zero. Large values of (13) suggest misspecification of the model.

Let $T_i(\theta_1)$ be the vector containing the elements of $\sum_{j=0}^{k}(1_i^{(j)}/P_i^{(j)})$ $(\text{vech}\partial^2 P_i^{(j)}/\partial\theta_1\partial\theta_1')$ included in the test and let $T_n(\theta_1) = n^{-1}\sum_{i=1}^{n} T_i(\theta_1)$. As shown by Lancaster (1984), for example, the IM test statistic is

$$nT_n(\hat{\theta}_1)'\hat{V}^{-1}T_n(\hat{\theta}_1) \xrightarrow{d} \chi_q^2,$$

where q is the number of elements in $T_i(\theta_1)$ and $\hat{V}$ is a consistent estimate of V, the asymptotic covariance matrix of $n^{\frac{1}{2}} T_n(\hat{\theta}_1)$. Under H_0,

$$V = n^{-1} \sum_{i=1}^{n} E\left[T_i(\theta_1) T_i(\theta_1)'\right]\Big|_{\theta_1^*} - n^{-1} \sum_{i=1}^{n} E\left[T_i(\theta_1) \frac{\partial \ell_i(\theta_1)}{\partial \theta_1'}\right]\Big|_{\theta_1^*} \mathbf{I}_{11}^{-1} \; n^{-1} \sum_{i=1}^{n} E\left[\frac{\partial \ell_i(\theta_1)}{\partial \theta_1} T_i(\theta_1)'\right]\Big|_{\theta_1^*}, \tag{15}$$

where $\ell_i(\theta_1) = \sum_{j=0}^{k} 1_i^{(j)} \log P_i^{(j)}$ is the i^{th} term in first sum in the log-likelihood function (4). $\hat{V}$ is obtained by replacing the expectations by sample averages evaluated at $\hat{\theta}_1$. The estimate of $n^{-1} \sum_{i=1}^{n} E[T_i(\theta_1) T_i(\theta_1)']|_{\theta_1^*}$, for example, is

$$n^{-1} \sum_{i=1}^{n} \sum_{j=0}^{k} \frac{1}{P_i^{(j)}} \left\{ \operatorname{vech} \frac{\partial^2 P_i^{(j)}}{\partial \theta_1 \partial \theta_1'} \right\} \left\{ \operatorname{vech} \frac{\partial^2 P_i^{(j)}}{\partial \theta_1 \partial \theta_1'} \right\}' \Bigg|_{\hat{\theta}_1}. \tag{16}$$

Finally, note that, since $\partial^2 P_i^{(j)} / \partial \gamma_k^2 = -\partial^2 P_i^{(j)} / \partial \gamma_k \partial \beta_1$, where β_1 is the intercept, some of the elements of (13) corresponding to the thresholds, γ_j, must be omitted from the test to avoid exact multicollinearity.

4. Ordered Probit Model.

In the OP model, the distribution of the errors is based on the normal:

$$F(u) = \Phi(u) \equiv \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} exp\{-\frac{t^2}{2}\} dt.$$

The distributional alternative is that ϵ_i / σ_i^* is a member of the Pearson family. Following Bera et al. (1984), let

$$q(t) = \int \frac{c_1 - t}{c_0 - c_1 t + c_2 t^2} \, dt, \tag{17}$$

where c_0, c_1, and c_2 are constants. Then the general density function is given by

$$f_c(u) = \exp\{q(u)\} / \int_{-\infty}^{\infty} \exp\{q(t)\} dt,$$

and F_c represents the corresponding distribution function. The constants c_0, c_1, and c_2 can be expressed in terms of the moments of the distribution of ϵ_i/σ_i^* as

$$c_0 = (4\mu_4 - 3\mu_3^2)(10\mu_4 - 12\mu_3^2 - 18)^{-1},$$

$$c_1 = \mu_3(\mu_4 + 3)(10\mu_4 - 12\mu_3^2 - 18)^{-1},$$

and

$$c_2 = (2\mu_4 - 3\mu_3^2 - 6)(10\mu_4 - 12\mu_3^2 - 18)^{-1},$$

where μ_3 and μ_4 are the third and fourth moments of the distribution (e.g., Johnson and Kotz 1970). Normality is implied by $c_1 = c_2 = 0$.[4] Note also that since the ϵ_i/σ_i^* are identically distributed under the alternative, any heterogeneity in the distribution of ϵ_i enters through the variance.

The null hypothesis to be tested is given by $H_0 : \theta_2 = 0$, where $\theta_2 = (b', a', c_1, c_2)'$; and under H_0, the model reduces to the OP model. More selective tests can be obtained by including only some of the elements of θ_2. A test for normality within the Pearson family, for example, is obtained by setting $\theta_2 = (c_1, c_2)'$.

From (11),

$$\frac{\partial L_c}{\partial \theta} = \sum_{i=1}^{n} \sum_{j=0}^{k} \frac{1_i^{(j)}}{F_{ci}^{(j+1)} - F_{ci}^{(j)}} \left(\frac{\partial F_{ci}^{(j+1)}}{\partial \theta} - \frac{\partial F_{ci}^{(j)}}{\partial \theta} \right). \tag{18}$$

Under H_0, $F_{ci}^{(j)}$ reduces to $\Phi_i^{(j)} \equiv \Phi\left((\gamma_j - x_i'\beta)/\sigma_i\right)$, $\sigma_i = h(\alpha_0^* + z_i'\alpha)^{\frac{1}{2}}$, and

[4] Substitute $c_1 = c_2 = 0$ and $c_0 = 1$ into equation (17), and integrate.

the derivatives of $F_{ci}^{(j)}$ with respect to the elements of θ are given by

$$\frac{\partial F_{ci}^{(j)}}{\partial \beta} = -\phi_i^{(j)} \frac{x_i}{\sigma_i}, \tag{19}$$

$$\frac{\partial F_{ci}^{(j)}}{\partial b} = -\phi_i^{(j)} \frac{w_i}{\sigma_i}, \tag{20}$$

$$\frac{\partial F_{ci}^{(j)}}{\partial \alpha} = -\phi_i^{(j)} \left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right) \frac{\partial \sigma_i}{\partial z_i'\alpha} \frac{z_i}{\sigma_i}, \tag{21}$$

$$\frac{\partial F_{ci}^{(j)}}{\partial a} = -\phi_i^{(j)} \left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right) \frac{\partial \sigma_i}{\partial \nu_i' a} \frac{\nu_i}{\sigma_i}, \tag{22}$$

$$\frac{\partial F_{ci}^{(j)}}{\partial \gamma_m} = \phi_i^{(j)} \frac{1(m=j)}{\sigma_i}, \quad m = 2, \ldots, k, \tag{23}$$

$$\frac{\partial F_{ci}^{(j)}}{\partial c_1} = \frac{1}{3}\phi_i^{(j)} \left[\left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right)^2 - 1\right], \tag{24}$$

and

$$\frac{\partial F_{ci}^{(j)}}{\partial c_2} = -\frac{1}{4}\phi_i^{(j)} \left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right) \left[\left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right)^2 + 3\right], \tag{25}$$

where $\phi_i^{(j)} = \phi\left((\gamma_j - x_i'\beta)/\sigma_i\right)$ and ϕ is the density function of the standard normal. Substituting (18) and (19)-(25) into (12) gives the form of the LM test. The expression for Σ is given by $\Sigma = n^{-1} \sum_{i=1}^{n} E\left[\bar{F}_{ci} I_i \bar{F}_{ci}'\right]\big|_{(\theta_1^{*\prime},0)'}$, where $\bar{F}_{ci}$ is equivalent to $\bar{P}_i$ in (7) with $F_{ci}^{(j)}$ and θ in place of $\Phi_i^{(j)}$ and θ_1, respectively. Note also that (19)-(23) represent the derivatives in the basic OP model.

In the probit model, $k = 1$, $1_i^{(1)} = y_i$, $P_i^{(1)} = \Phi(x_i'\beta/\sigma_i) \equiv \Phi_i$, $\phi_i^{(1)} = \phi(x_i'\beta/\sigma_i) \equiv \phi_i$, and (18) reduces to

$$\frac{\partial L_c}{\partial \beta} = \sum_{i=1}^{n} \frac{y_i - \Phi_i}{\Phi_i(1-\Phi_i)} \phi_i \frac{x_i}{\sigma_i},$$

$$\frac{\partial L_c}{\partial \alpha} = -\sum_{i=1}^{n} \frac{y_i - \Phi_i}{\Phi_i(1-\Phi_i)} \phi_i \frac{x_i'\beta}{\sigma_i} \frac{\partial \sigma_i}{\partial z_i'\alpha} \frac{z_i}{\sigma_i},$$

$$\frac{\partial L_c}{\partial b} = \sum_{i=1}^{n} \frac{y_i - \Phi_i}{\Phi_i(1-\Phi_i)} \phi_i \frac{w_i}{\sigma_i},$$

$$\frac{\partial L_c}{\partial a} = -\sum_{i=1}^{n} \frac{y_i - \Phi_i}{\Phi_i(1-\Phi_i)} \phi_i \frac{x_i'\beta}{\sigma_i} \frac{\partial \sigma_i}{\partial \nu_i' a} \frac{\nu_i}{\sigma_i},$$

$$\frac{\partial L_c}{\partial c_1} = -\frac{1}{3} \sum_{i=1}^{n} \frac{y_i - \Phi_i}{\Phi_i(1-\Phi_i)} \phi_i \left[\left(\frac{x_i'\beta}{\sigma_i} \right)^2 - 1 \right],$$

and

$$\frac{\partial L_c}{\partial c_2} = \frac{1}{4} \sum_{i=1}^{n} \frac{y_i - \Phi_i}{\Phi_i(1-\Phi_i)} \phi_i \left(\frac{x_i'\beta}{\sigma_i} \right) \left[\left(\frac{x_i'\beta}{\sigma_i} \right)^2 + 3 \right].$$

One interpretation of the LM test is obtained by rewriting the elements of the score vector in (18). Under H_0, evaluated at θ_1^*, and using (20),

$$\begin{aligned} &\sum_{j=0}^{k} \frac{1_i^{(j)}}{F_{ci}^{(j+1)} - F_{ci}^{(j)}} \left(\frac{\partial F_{ci}^{(j+1)}}{\partial b} - \frac{\partial F_{ci}^{(j)}}{\partial b} \right) \\ &= \sum_{j=0}^{k} \frac{-1_i^{(j)}}{\Phi_i^{(j+1)} - \Phi_i^{(j)}} \left(\phi_i^{(j+1)} - \phi_i^{(j)} \right) \frac{w_i}{\sigma_i^*} \\ &= \sum_{j=0}^{k} \frac{1_i^{(j)}}{\Phi_i^{(j+1)} - \Phi_i^{(j)}} \int_{\frac{\gamma_j^* - x_i'\beta^*}{\sigma_i^*}}^{\frac{\gamma_{j+1}^* - x_i'\beta^*}{\sigma_i^*}} u\phi(u)du \frac{w_i}{\sigma_i^*} \\ &= \sum_{j=0}^{k} 1_i^{(j)} E\left[\epsilon_i/\sigma_i^* | x_i, z_i, y_i = j\right] \frac{w_i}{\sigma_i^*}. \end{aligned} \tag{26}$$

That is, the LM test is based on the correlation between ϵ_i and the omitted variables w_i. But ϵ_i is not observed, so instead the test uses the expected value of ϵ_i given y_i, $\sum_{j=0}^{k} 1_i^{(j)} E\left[\epsilon_i | x_i, z_i, y_i = j\right]$. These expected values are known as the "generalized residuals" (Gourieroux et al. 1987). In the probit model, they reduce to the $(y_i - \Phi_i)\phi_i/(\Phi_i(1-\Phi_i))$ appearing in the expressions for $\partial L_c/\partial\beta$, etc. Similarly, the test for omitted variables in the variance equation is based on the correlation between $(\epsilon_i/\sigma_i^*)^2$ and ν_i, and the test for normality is based on the third and fourth moments of ϵ_i/σ_i^*: From (22), (24), and (25), under H_0,

$$\frac{1}{F_{ci}^{(j+1)} - F_{ci}^{(j)}} \left(\frac{\partial F_{ci}^{(j+1)}}{\partial a} - \frac{\partial F_{ci}^{(j)}}{\partial a} \right)$$
$$= E\left[\left(\epsilon_i/\sigma_i^*\right)^2 - 1 | x_i, z_i, y_i = j \right] \frac{\partial \sigma_i}{\partial \nu_i' a} \frac{\nu_i}{\sigma_i^*}, \tag{27}$$
$$\frac{1}{F_{ci}^{(j+1)} - F_{ci}^{(j)}} \left(\frac{\partial F_{ci}^{(j+1)}}{\partial c_1} - \frac{\partial F_{ci}^{(j)}}{\partial c_1} \right)$$
$$= -\frac{1}{3} E\left[\left(\epsilon_i/\sigma_i^*\right)^3 - 3\left(\epsilon_i/\sigma_i^*\right) | x_i, z_i, y_i = j \right], \tag{28}$$

and

$$\frac{1}{F_{ci}^{(j+1)} - F_{ci}^{(j)}} \left[\frac{\partial F_{ci}^{(j+1)}}{\partial c_2} - \frac{\partial F_{ci}^{(j)}}{\partial c_2} \right] = \frac{1}{4} E\left[\left(\epsilon_i/\sigma_i^*\right)^4 - 3 | x_i, z_i, y_i = j \right]. \tag{29}$$

Note also that when the null hypothesis implies a homoskedastic error, the derivative with respect to α does not appear and $\partial \sigma_i / \partial \nu_i' a$ reduces to a constant. As in tests for heteroskedasticity in other models, the functional form of h is irrelevant.

The first elements of the IM test are those corresponding to β. Differentiating $P_i^{(j)}$ gives

$$\frac{\partial^2 P_i^{(j)}}{\partial \beta \partial \beta'} = -\phi_i^{(j+1)} \left(\frac{\gamma_{j+1} - x_i'\beta}{\sigma_i} \right) \frac{x_i x_i'}{\sigma_i^{*2}} + \phi_i^{(j)} \left(\frac{\gamma_j - x_i'\beta}{\sigma_i} \right) \frac{x_i x_i'}{\sigma_i^{*2}},$$

and hence, arguing as in equation (26),

$$\left. \frac{1}{P_i^{(j)}} \frac{\partial^2 P_i^{(j)}}{\partial \beta \partial \beta'} \right|_{\theta_i^*} = E\left[(\epsilon_i/\sigma_i^*)^2 - 1 | x_i, z_i, y_i = j \right] \frac{x_i x_i'}{\sigma_i^{*2}}. \tag{30}$$

These elements of the IM test are based on the squared residuals and in this sense are "testing for neglected heterogeneity" (Chesher 1984). The IM test differs from the LM test in that it multiplies the generalized residuals by different weights – equation (30) uses $x_i x_i'/\sigma_i^*$, whereas (27) uses the (presumably) more carefully chosen weights based on ν_i.

Applying the same arguments to the other terms in the IM test gives

$$\frac{1}{P_i^{(j)}}\frac{\partial^2 P_i^{(j)}}{\partial\beta\partial\alpha} = E\left[(\epsilon_i/\sigma_i^*)^3 - 3(\epsilon_i/\sigma_i^*)|x_i, z_i, y_i = j\right]\frac{x_i z_i'}{\sigma_i^*}, \tag{31}$$

$$\frac{1}{P_i^{(j)}}\frac{\partial^2 P_i^{(j)}}{\partial\alpha\partial\alpha'} = E\left[\left((\epsilon_i/\sigma_i^*)^4 - 3\right) - 4\left((\epsilon_i/\sigma_i^*)^2 - 1\right)|x_i, z_i, y_i = j\right] z_i z_i', \tag{32}$$

$$\begin{aligned}
&\frac{\partial^2 P_i^{(j)}}{\partial\gamma_m\partial\gamma_\ell} \\
&= -1(m = \ell = j+1)P(y_i \le j+1)E\left[\left((\epsilon_i/\sigma_i^*)^2 - 1\right)|x_i, z_i, y_i \le j+1\right]\frac{1}{\sigma_i^{*2}} \\
&+ 1(m = \ell = j)P(y_i \le j)E\left[\left((\epsilon_i/\sigma_i^*)^2 - 1\right)|x_i, z_i, y_i \le j\right]\frac{1}{\sigma_i^{*2}},
\end{aligned} \tag{33}$$

$$\begin{aligned}
&\frac{\partial^2 P_i^{(j)}}{\partial\gamma_m\partial\beta} \\
&= 1(m = j+1)P(y_i \le j+1)E\left[\left((\epsilon_i/\sigma_i^*)^2 - 1\right)|x_i, z_i, y_i \le j+1\right]\frac{x_i}{\sigma_i^{*2}} \\
&- 1(m = j)P(y_i \le j)E\left[\left((\epsilon_i/\sigma_i^*)^2 - 1\right)|x_i, z_i, y_i \le j\right]\frac{x_i}{\sigma_i^{*2}},
\end{aligned} \tag{34}$$

and

$$\begin{aligned}
&\frac{\partial^2 P_i^{(j)}}{\partial\gamma_m\partial\alpha} = \\
&- 1(m = j+1)P(y_i \le j+1)E\left[\left((\epsilon_i/\sigma_i^*)^3 - 3(\epsilon_i/\sigma_i^*)\right)|x_i, z_i, y_i \le j+1\right]\frac{z_i}{\sigma_i^*} \\
&+ 1(m = j)P(y_i \le j)E\left[\left((\epsilon_i/\sigma_i^*)^3 - 3(\epsilon_i/\sigma_i^*)\right)|x_i, z_i, y_i \le j\right]\frac{z_i}{\sigma_i^*}.
\end{aligned} \tag{35}$$

Notice that because z_i does not include an intercept, the IM test does not perform the simple test comparing the third moment to zero and the fourth moment to three. Similarly, if there is no heteroskedasticity in the specified model ($\alpha = 0$), then the third and fourth moment terms do not appear at all. In this case, normality is being tested indirectly through (30), (33), and (34) as in equation (14). (33) - (35) also show that many of the elements of (13) are equal to zero. These elements are omitted from the test.

5. Ordered Logit Model.

A simple generalization of the logistic distribution is that considered by Moon (1988):

$$F_c(u) = \frac{1}{(1 + exp\{-u\})^{1+c}} \qquad c > -1. \tag{36}$$

This generalization relaxes the symmetry of the logistic distribution, with skewness the same sign as c, and has most affect on the kurtosis when $c < 0$ due to a long tail to the left.[5] In the LM test, the derivatives of $F_{ci}^{(j)}$ with respect to β, b, α, a, and the γ_m are given by (19)-(23) with $\phi_i^{(j)}$ replaced by $\Lambda_i^{(j)}(1 - \Lambda_i^{(j)})$, where $\Lambda_i^{(j)} = \Lambda\left((\gamma_j - x_i'\beta)/\sigma_i\right)$. The derivative of $F_c(z)$ with respect to c is given by $(1+c)^{-1}F_c(z)logF_c(z)$. The score vector is defined as in (18) and the LM test as in (12).

The interpretation of the test in terms of residuals is obtained by directly from the score vector. Write

$$\frac{\partial L_c}{\partial \theta} = \sum_{i=1}^{n} \sum_{j=0}^{k} \frac{1_i^{(j)} - P_i^{(j)}}{P_i^{(j)}} \frac{\partial P_i^{(j)}}{\partial \theta_1}, \tag{37}$$

and the residuals are the $1_i^{(j)} - P_i^{(j)}$.[6] In the logit model, $k = 1$, $y_i = 1_i^{(1)}$, and (37) reduces to

$$\begin{aligned} \frac{\partial L_c}{\partial \beta} &= \sum_{i=1}^{n} (y_i - P_i^{(1)}) \frac{x_i}{\sigma_i}, \\ \frac{\partial L_c}{\partial \alpha} &= -\sum_{i=1}^{n} (y_i - P_i^{(1)}) \frac{x_i'\beta}{\sigma_i} \frac{\partial \sigma_i}{\partial z_i'\alpha} \frac{z_i}{\sigma_i}, \\ \frac{\partial L_c}{\partial b} &= \sum_{i=1}^{n} (y_i - P_i^{(1)}) \frac{w_i}{\sigma_i}, \end{aligned} \tag{38}$$

[5] Moon (1988) gives the moments and cumulants of F_c.

[6] Gourieroux et al. (1987) base their definition of generalized residuals on the exponential model – the latent variable y_i^* is exponential and the residual is $E[y_i^* - E[y_i^*|x_i]|y_i]$. In the logit model, y_i itself is exponential and $E[y_i - E[y_i^*|x_i]|y_i] = y_i - E[y_i|x_i]$.

$$\frac{\partial L_c}{\partial a} = -\sum_{i=1}^{n}(y_i - P_i^{(1)})\frac{x_i'\beta}{\sigma_i}\frac{\partial \sigma_i}{\partial \nu_i' a}\frac{\nu_i}{\sigma_i},$$

and

$$\frac{\partial L_c}{\partial c} = \sum_{i=1}^{n}\frac{y_i - P_i^{(1)}}{1 - P_i^{(1)}} log P_i^{(1)},$$

where $P_i^{(1)} = \Lambda(x_i'\beta)$. The test is for correlation between the residuals and the extra terms.

An alternative to the generalized logistic distribution in (36) is

$$F_c(u) = \frac{1}{1 + exp\{-(u + c_1 u^2 + c_2 u^3)\}} \tag{39}$$

(e.g., Ruud 1984, Pagan and Vella 1989).[7] Corresponding to $\partial F_{ci}^{(j)}/\partial c_1$ and $\partial F_{ci}^{(j)}/\partial c_2$ in the OP model are

$$\frac{\partial F_{ci}^{(j)}}{\partial c_1} = F_{ci}^{(j)}(1 - F_{ci}^{(j)})\left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right)^2 \tag{40}$$

and

$$\frac{\partial F_{ci}^{(j)}}{\partial c_2} = F_{ci}^{(j)}(1 - F_{ci}^{(j)})\left(\frac{\gamma_j - x_i'\beta}{\sigma_i}\right)^3. \tag{41}$$

Turning to the IM test, I simply note it is straightforward to derive the form of the test. Furthermore, the logistic distribution does not satisfy the conditions in Chesher (1984) and the test cannot be interpreted as a test for heteroskedasticity. Hence, a detailed analysis of the test is omitted.

6. Local Alternatives.

Both the LM and IM tests are examples of generalized method of moments tests. Newey (1985) has studied the power properties of such tests

[7] Of course, the same type of alternative could have been used in the OP model, i.e., $F(u) = \Phi(u + c_1 u^2 + c_3 u^3)$. The resulting test would be quite similar to the LM test based on (24) and (25).

under sequences of local alternatives to the null hypothesis. In this section, I summarize the relevant behavior of the LM and IM tests and show how these results were used in the design of the Monte Carlo experiment.

I assume that the observed data, $d_i = (y_i, x_i', w_i', z_i', \nu_i')', i = 1, \ldots, n$, is drawn from a probability density function $f(d|\theta_1^*, \lambda_n)$, where $\lambda_n = \lambda_0 + \tau/\sqrt{n}$ and f satisfies the conditions in Newey (1985).[8] The vector λ represents parameters that affect the correctness of the specification and at λ_0, the model is correctly specified. For example, if $\lambda = \theta_2$ and $\lambda_0 = 0$, the misspecification corresponds to the alternative in the LM test.

Denote the moment conditions to be tested by $m(d_i, \theta_1)$. In the LM test $m(d_i, \theta_1)$ is obtained from $\partial L_c/\partial\theta_2$ and in the IM test $m(d_i, \theta_1)$ is obtained from (13). At λ_0, $E[m(d_i, \theta_1)]|_{\theta_1^*} = 0$ and $E[\partial\ell_i(\theta_1)/\partial\theta_1]|_{\theta_1^*} = 0$; and by a mean value expansion about the true parameter vector θ_1^*,

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} m(d_i, \hat{\theta}_1) = n^{-\frac{1}{2}} \sum_{i=1}^{n} m(d_i, \theta_1^*) + n^{-1} \sum_{i=1}^{n} \frac{\partial m(d_i, \theta_1^{\#})}{\partial\theta_1'} n^{\frac{1}{2}}(\hat{\theta}_1 - \theta_1^*)$$
$$= \left[-\sum_{i=1}^{n} \frac{\partial m(d_i, \theta_1^{\#})}{\partial\theta_1'} \frac{\partial^2 L(\theta^{\#\#})}{\partial\theta_1 \partial\theta_1}^{-1} : I \right] \left[n^{-\frac{1}{2}} \sum_{i=1}^{n} g_i(\theta_1^*) \right],$$

where $g_i(\theta_1) = (\partial\ell_i(\theta_1)/\partial\theta_1', m(d_i, \theta_1)')'$ and $\theta_1^{\#}$ and $\theta_1^{\#\#}$ are mean values. But

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} g_i(\theta^*) \xrightarrow{d} N(K\tau, W),$$

where $K = E_0\,[g_i(\theta_1)\partial \log f(d_i|\theta_1^*, \lambda)/\partial\lambda']$, $W = E_0\,[g_i(\theta_1)g_i(\theta_1)']$, and E_0 means the expectation is evaluated at θ_1^* and λ_0. It follows that

$$n^{-1} \sum_{i=1}^{n} m_i(d_i, \hat{\theta}_1)'\ \hat{Q}_n^{-1} \sum_{i=1}^{n} m_i(d_i, \hat{\theta}_1) \xrightarrow{d} \chi_r^2(\mu),$$

[8] Newey refers to f as a probability density function (h in his notation), but his analysis applies to models on categorical data. In fact, he uses the probit as an example.

where r is the number of elements in $m_i(d_i, \theta_1)$, $\chi_r^2(\mu)$ is a noncentral χ_r^2 with noncentrality parameter $\mu = \tau' K' [M\mathbf{I}_{11}^{-1} : I]' Q^{-1} [M\mathbf{I}_{11}^{-1} : I] K\tau$, $M = -E_0[\partial m(d_i, \theta_1)/\partial\theta_1']$, and $\hat{Q}_n$ is an estimate of $Q = [M\mathbf{I}_{11}^{-1} : I] W [M\mathbf{I}_{11}^{-1} : I]'$.

In the case of the LM test, if $\lambda = \theta_2$, then $m(d_i, \theta_1) = \partial\ell_{ci}/\partial\theta_2$, $M = -\Sigma_{21}$, $W = \Sigma$, $Q = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, $\partial \log f(d_i|\theta_1, \lambda)/\partial\lambda = \partial\ell_{ci}/\partial\theta_2$, and $K = (\Sigma_{12}' : \Sigma_{22}')'$, where ℓ_{ci} is the contribution of observation i to L_c in (11). The noncentrality parameter becomes

$$\mu = \tau' Q \tau. \tag{42}$$

When the sample is large and the alternative is close to the null hypothesis, (42) can be used to approximate the behavior of the test statistics. For example, suppose the test is for a single moment condition ($r = 1$) and has size 0.05 asymptotically. The test statistic will be asymptotically distributed as $\chi_1^2(\mu)$, so that if $\sqrt{\mu} = 1.96$, the square root of the test statistic will be appproximately $N(1.96, 1)$ and the null will be rejected approximately 50% of the time. In other words, the test will have power approximately equal to 0.5. If the test is an LM test and the alternative being tested actually generated the data, the value of θ_2 can be determined from τ via equation (42). This calculation was used in the Monte Carlo experiment below to determine parameter values under the alternative, and therefore to see how well the small sample properties of test statistics are approximated by the asymptotic properties.

7. Monte Carlo Experiment.

7.1. Design. The data generation process is given by

$$y_i^* = \beta_1^* + \beta_2^* x_{1i} + b^* w_i + \epsilon_i \tag{43}$$

and

$$\sigma_i^* = exp\{\nu_i a^*\}, \tag{44}$$

where x_{1i} and w_i are bivariate normal with means zero, variances one, and correlation ρ, ν_i is the time trend $0.10 + 0.01i$, and ϵ_i is standard normal or logistic under the null and χ^2 or t (normalized to have mean zero and variance σ_0^2) under the alternative.[9] β_2^* is set to $2\sigma_0$, but β_1^* and the γ_i^* vary across experiments. Under the null hypothesis, $b^* = a^* = c = 0$. The number of observations is 100, 300, or 500; and the number of replications is 1000.

I include the following LM tests in the experiments:

i) tests for the separate inclusion of w_i and ν_i (denoted by LM-1 and LM-2, respectively);

ii) tests for non-normal/non-logistic errors (using the Pearson and denoted by LM-3 in the OP model, and using (39) or (36) and denoted by LM-3 and LM-5, respectively, in the OL model); and

iii) joint tests for w_i, ν_i, and non-normal/non-logistic errors (denoted by LM-4 in the OP model, and by LM-4 and LM-6 in the OL model, using (39) and (36), respectively).

The IM tests are those based on including in $T_i(\theta_1)$:

i) the element of (30) corresponding to the second derivative with respect to β_2 (IM-1); and

ii) this element plus the elements from (34) (IM-2).

[9] Strictly speaking, the time trend is nonstochastic and does not satisfy the conditions on the variables imposed in Bera et al. (1984) or Newey (1985). I use it because it implies a very distinct form of heteroskedasticity. It was also used in Davidson and MacKinnon (1984).

I also include two versions of the chi-squared goodness of fit tests described in Andrews (1988) (χ^2-1 and χ^2-2). χ^2-1 compares the actual and estimated frequencies in the categories for y_i, and in χ^2-2 the observations are also split according to whether $x_i'\hat{\beta}$ is above or below median$(x_i'\hat{\beta})$. χ^2-1 is not defined in the logit model because, by the first-order condition with respect to β, the difference between the actual and estimated frequencies is zero (see equation 38). The covariance matrices of the vectors of (actual-estimated) frequencies are calculated using the estimate $\hat{\Sigma}_1$ in Andrews' equation (3.14), except that the information matrix is estimated using the scoring estimate based on the third term in (8), rather than his V_n.

I use different realizations of x_{1i} and w_i in different experiments, but the same realization across replications within each experiment. If a realization of the ϵ_i gives a category of y_i with zero frequency, the realization is discarded.

7.2. Size. The first two experiments consider the ordinary probit and logit models and the results are presented in Tables I and II. The tables give, for each test, the degrees of freedom, the mean of the test statistic across replications, the percentages of rejections of the null hypothesis at the nominal 5% and 10% levels, and the estimated critical values for the test to have sizes 5% and 10%. I also calculated the t-statistics for the hypotheses that the means equal the degrees of freedom and the percentages of rejections equal the nominal sizes. A * implies that the t-statistic was less that -2.0 – the mean or percentage was too small – while a @ implies that the t-statistics was greater than 2.0 – the mean or percentage was too big. The results show that in the probit model with 100 observations, many of the tests reject less than 5/10% of the time. This behavior continues beyond 100 observations

TABLE I

Performance of Tests under Null - Probit Model

Test:	LM-1	LM-2	LM-3	LM-4	χ^2-1	χ^2-2	IM-1	IM-2
df:	1	1	2	4	1	2	1	1
Means:								
100 obs	0.99	0.89*	1.43*	3.34*	0.99	1.92	0.61*	0.61*
300 obs	1.08	0.95	1.91	3.96	0.97	1.97	0.96	0.96
500 obs	1.02	1.00	1.91	3.94	0.99	1.99	0.95	0.95
% rejections (5%):								
100 obs	4.2	2.8*	3.2*	3.4*	4.9	3.9	2.0*	2.0*
300 obs	6.0	4.6	4.4	4.4	5.2	5.0	3.1*	3.1*
500 obs	4.4	5.1	4.6	5.9	5.2	4.8	3.5*	3.5*
% rejections (10%):								
100 obs	9.8	8.2	4.2*	6.5*	9.6	7.7*	2.9*	2.9*
300 obs	11.3	9.1	6.1*	8.7	9.1	9.6	4.9*	4.9*
500 obs	10.3	9.8	7.6*	9.5	10.5	10.1	7.1*	7.1*
estimated critical values (5%):								
100 obs	3.53	3.27	4.03	8.31	3.79	5.51	1.72	1.72
300 obs	3.99	3.77	5.63	9.29	3.88	5.99	2.60	2.60
500 obs	3.69	3.94	5.76	9.91	3.87	5.91	3.36	3.36
estimated critical values (10%):								
100 obs	2.67	2.44	2.49	6.39	2.65	4.17	1.20	1.20
300 obs	2.93	2.63	3.28	7.39	2.51	4.51	1.69	1.69
500 obs	2.81	2.69	3.76	7.69	2.79	4.62	2.21	2.21

Note: * means the asymptotic t-statistic for mean = degrees of freedom or % rejections = 5/10 was less than -2.0.

in the IM tests. Otherwise the tests behave reasonably well. In the logit model the IM tests again rejects too infrequently, although the other tests seem to reject more often.

Next, Tables III and IV give the results for specific OP and OL models. These are based on models with four categories for y_i and parameter values such that the expected frequencies in the four categories are 0.1, 0.5, 0.3, and 0.1. $((\beta_0^*, \gamma_2^*, \gamma_3^*) = (5.18, 6.20, 10.36)$ in the OL model and $(2.865, 3.43,$

TABLE II

Performance of Tests under Null - Logit Model

Test:	LM-1	LM-2	LM-3	LM-4	LM-5	LM-6	χ^2-1	χ^2-2	IM-1	IM-2
df:	1	1	2	4	1	3	0	1	1	1
Means:										
100 obs	1.08	0.96	1.90	3.94	1.05	3.09	-	1.07	1.00	1.00
300 obs	1.11$^{@}$	1.05	1.81	3.98	1.03	3.20$^{@}$	-	1.00	0.90$^{@}$	0.90$^{@}$
500 obs	1.05	1.01	1.85	3.92	1.00	3.06	-	0.95	0.85$^{@}$	0.85$^{@}$
% rejections (5%):										
100 obs	6.6$^{@}$	4.6	6.7$^{@}$	6.7$^{@}$	5.6	5.6	-	5.9	4.8	4.8
300 obs	5.8	5.4	4.9	6.5$^{@}$	6.0	6.4$^{@}$	-	5.0	3.5*	3.5*
500 obs	6.0	5.0	4.9	5.6	4.7	5.4	-	5.0	2.7*	2.7*
% rejections (10%):										
100 obs	12.4$^{@}$	9.7	8.4	9.7	9.6	11.9$^{@}$	-	11.7	7.6*	7.6*
300 obs	10.8	9.9	6.6*	10.3	10.7	11.4	-	11.0	6.4*	6.4*
500 obs	11.3	9.6	6.7*	8.9	10.4	10.4	-	9.3	6.5*	6.5*
estimated critical values (5%):										
100 obs	4.24	3.67	7.37	10.86	4.06	8.10	-	4.00	3.69	3.69
300 obs	4.02	4.00	5.59	10.25	4.36	8.64	-	3.83	3.03	3.03
500 obs	4.19	3.89	5.93	10.01	3.71	8.09	-	3.86	2.96	2.96
estimated critical values (10%):										
100 obs	3.12	2.67	3.91	7.76	2.58	6.62	-	2.87	2.01	2.01
300 obs	2.81	2.69	3.51	8.00	2.92	6.71	-	2.89	2.16	2.16
500 obs	2.92	2.66	3.56	7.38	2.73	6.35	-	2.63	2.21	2.21

Note: * means the asymptotic t-statistic for mean = degrees of freedom or % rejections = 5/10 was less than -2.0. $^{@}$ means the t-statistic exceeded 2.0.

5.73) in the OP model). χ^2-2 rejects too often, although since the mean is often below the degrees of freedom, this suggests that the distribution is more disperse than the asymptotic distribution. Generally, the means and rejection percentages are closer to the nominal values in the OL model.

I also investigated a number of other models with different thresholds and values of β_1^* and ρ. These gave comparable results. More importantly, I investigated OPG versions of the tests. As noted above, the tests had poor small sample properties and the results are not reported.

TABLE III

Performance of Tests under Null - Ordered Probit Model

Test:	LM-1	LM-2	LM-3	LM-4	χ^2-1	χ^2-2	IM-1	IM-2
df:	1	1	2	4	3	6	1	5
Means:								
100 obs	1.05	0.91*	1.47*	3.42*	2.66*	4.70*	0.81*	4.23*
300 obs	1.01	0.90*	1.95	3.86	3.00	6.42	0.93	4.71*
500 obs	1.02	0.97	1.84	3.84	2.82*	5.39*	0.95	4.64*
% rejections (5%):								
100 obs	5.4	3.6*	3.0*	3.4*	4.2	3.2*	2.7*	4.5
300 obs	5.3	4.0	5.8	5.6	6.0	9.1°	3.9	6.0
500 obs	4.2	3.7	4.4	4.4	3.8	7.2°	3.8	5.6
% rejections (10%):								
100 obs	10.5	8.3	4.2*	6.8*	8.1*	5.0*	4.9*	7.7*
300 obs	10.2	8.2	7.8*	9.4	10.5	11.2	7.2*	8.8
500 obs	10.1	9.8	7.7*	8.0*	8.7	9.3	8.2	8.0*
estimated critical values (5%):								
100 obs	4.04	3.52	4.10	8.24	7.39	10.73	2.58	10.82
300 obs	3.96	3.45	6.40	10.06	8.05	18.42	3.23	11.92
500 obs	3.66	3.58	5.74	9.00	7.20	15.66	3.40	11.31
estimated critical values (10%):								
100 obs	2.77	2.42	2.60	6.41	5.70	8.05	1.91	7.95
300 obs	2.77	2.36	3.88	7.57	6.35	11.68	2.23	8.83
500 obs	2.72	2.67	4.04	7.15	5.95	10.19	2.45	8.62

Note: * means the asymptotic t-statistic for mean = degrees of freedom or % rejections = 5/10 was less than -2.0. ° means the t-statistic exceeded 2.0.

7.3. Power. The basic models are the same as those used in Tables I-IV; and I experiment with the following alternatives: an omitted variable in the regression equation ($b^* \neq 0$), an omitted variable in the variance equation ($a^* \neq 0$), and non-normal/non-logistic errors with the ϵ_i either χ^2 or t. The values of b^* are obtained from (42) so that the approximate asymptotic rejection probability in LM-1 is either 0.25, 0.50, or 0.75. For a^*, the same procedure is applied to LM-2.

TABLE IV

Performance of Tests under Null - Ordered Logit Model

Test:	LM-1	LM-2	LM-3	LM-4	LM-5	LM-6	χ^2-1	χ^2-2	IM-1	IM-2
df:	1	1	2	4	1	3	3	6	1	5
Means:										
100 obs	0.99	1.00	1.76*	3.75*	1.03	3.02	2.74	5.57	1.10@	5.35@
300 obs	0.97	0.99	1.85	3.82	1.01	2.97	2.79	5.74	0.98	4.92
500 obs	0.95	0.98	2.13	4.07	1.07	3.00	3.18	6.20	0.98	5.05
% rejections (5%):										
100 obs	4.5	5.1	5.4	5.2	5.2	4.5	7.1@	9.9@	5.7	6.5@
300 obs	4.8	5.0	4.1	4.5	4.6	4.6	6.3	8.2@	3.7	4.8
500 obs	4.1	4.9	6.2	6.2	5.9	4.9	6.7@	8.0@	5.3	4.9
% rejections (10%):										
100 obs	10.1	10.1	7.4*	8.1*	10.0	10.1	9.6	12.5@	11.0	11.0
300 obs	9.7	9.9	6.6*	8.3	10.0	9.6	8.8	11.4	8.7	9.4
500 obs	8.9	9.7	10.2	10.1	11.0	8.8	10.6	12.9@	9.9	9.8
estimated critical values (5%):										
100 obs	3.75	3.89	6.26	9.80	3.95	7.56	9.57	18.72	4.15	12.01
300 obs	3.79	3.87	5.53	8.98	3.80	7.56	9.26	15.30	3.48	10.99
500 obs	3.41	3.79	6.81	10.24	4.08	7.66	9.26	15.10	3.88	10.99
estimated critical values (10%):										
100 obs	2.75	2.71	3.56	7.01	2.70	6.26	6.03	12.49	2.86	9.71
300 obs	2.61	2.65	3.64	7.33	2.74	6.09	5.76	11.54	2.52	9.09
500 obs	2.51	2.69	4.71	7.83	2.86	6.08	6.43	11.58	2.70	9.13

Note: * means the asymptotic t-statistic for mean = degrees of freedom or % rejections = 5/10 was less than -2.0. @ means the t-statistic exceeded 2.0.

Table V details the behavior of LM-1 and LM-2 under the first two alternatives (labeled "Omitted Vble" for $b^* \neq 0$ and "Hetero" for $a^* \neq 0$). Shown are the values of b^* or a^*, the means of the test statistics, the percentages of rejections using the nominal 5% critical values, and the indicators of whether the means and rejection percentages are significantly different from their asymptotic values. (The asymptotic means of the test statistics are equal to the degrees of freedom plus μ: 2.65, 4.84, and 7.94 for prob-

TABLE V

Behavior of LM Tests for Omitted Variables and Heteroskedasticity under Local Alternatives.

	Obs	Omitted Vble/LM-1			Hetero/LM-2		
Probability:		0.25	0.50	0.75	0.25	0.50	0.75
a) Probit:							
b^*/a^*	100	0.241	0.368	0.495	0.790	1.205	1.620
Mean	100	2.60	4.41*	6.60*	2.35*	3.53*	4.56*
% rej	100	25.5	49.1	73.1	21.5*	38.0*	50.1*
b^*/a^*	500	0.101	0.154	0.207	0.070	0.107	0.144
Mean	500	2.40*	4.48*	7.45*	2.58	4.60	7.62
% rej	500	21.8*	46.9	74.0	24.9	48.8	73.5
b) Ordered Probit:							
b^*/a^*	100	0.176	0.269	0.361	0.502	0.765	1.029
Mean	100	2.61	4.90	7.70	2.80	5.48*	8.22
% rej	100	24.4	53.0	73.7	26.8	56.6	79.4$^{@}$
b^*/a^*	500	0.076	0.116	0.157	0.046	0.071	0.095
Mean	500	2.83	4.89	7.64	2.74	5.37*	9.42$^{@}$
% rej	500	27.5	50.6	73.9	26.3	53.8	83.1$^{@}$
c) Logit:							
b^*/a^*	100	0.367	0.560	0.753	0.906	1.344	1.859
Mean	100	2.74	4.81	7.59*	2.60	3.83*	5.49*
% rej	100	27.2	49.2	76.1	23.5	41.2*	58.5*
b^*/a^*	500	0.192	0.277	0.393	0.079	0.122	0.163
Mean	500	2.70	4.63	8.10	2.81	4.62	7.74
% rej	500	26.1	47.4	76.0	27.9	48.8	76.4
d) Ordered Logit:							
b^*/a^*	100	0.312	0.478	0.640	0.563	0.890	1.153
Mean	100	2.70	4.65	8.11	2.88	5.79$^{@}$	8.96$^{@}$
% rej	100	25.8	47.8	77.7	28.2	59.5	81.9
b^*/a^*	500	0.136	0.206	0.278	0.053	0.080	0.108
Mean	500	2.68	4.86	7.69	2.68	5.29$^{@}$	9.20$^{@}$
% rej	500	24.8	53.0	74.4	24.8	53.5	83.1

Note: * means the asymptotic t-statistic for mean = asymptotic mean or % rejections = asymptotic value was less than -2.0. $^{@}$ means the t-statistic exceeded 2.0.

abilities 0.25, 0.50, and 0.75, respectively.) While the means and rejection percentages are often significantly different from the asymptotic values (but less so in the OL model), the asymptotic values provide at least reasonable predictions of the behavior of the test statistics. Notice also that somewhat larger parameters values are required in the probit and logit than in the OP and OL. This reflects the fact that less information is available on the distribution of y_i^* (e.g., Skeels and Vella 1991).

For the χ^2 and t alternatives and in the OP model at least, the approximations based on (42) were much worse than for $b^* \neq 0$ and $a^* \neq 0$.[10] With degrees of freedom set to 5, 10, and 20, I found that using (42) led to systematic over-estimation of the rejection probability in the χ^2 and under-estimation in the t. Because the results add little to the discussion, they are omitted.

More detailed lists of rejection percentages are presented in Tables VI and VII. These are based on b^* and a^* parameter values implying the asymptotic rejection probabilities in LM-1 and LM-2 are 0.5, and on ten degrees of freedom in the χ^2 and t distributions. I simply set the degrees of freedom because using (42) leads to poor approximations. The fixed alternatives in the χ^2 and t distributions mean that, unlike the omitted variable and heteroskedasticity alternatives, the powers of the tests increase with the numbers of observations. The rejection percentages are also based on the estimated 5% critical values. Because of their higher degrees of freedom, the general LM tests (LM-4 and LM-6) have less power against specific alternatives than the LM tests designed for those alternatives. In the logit and OL models with the χ^2 alternative, LM-3 has lower power than LM-5 due to its higher de-

[10] Recall that the χ^2 and t are Type-III ($c_2 = 0$) and Type-VII ($c_1 = 0$) Pearson distributions, respectively.

TABLE VI

Powers in Probit and Ordered Probit Models
Using Estimated 5% Critical Values

Test:	LM-1	LM-2	LM-3	LM-4	χ^2-1	χ^2-2	IM-1	IM-2
a) Probit:								
df:	1	1	2	4	1	2	1	1
Omitted Variable:								
100 obs	52.8	5.1	4.5	27.5	3.8	4.4	3.8	3.8
500 obs	49.0	4.8	6.3	25.4	4.4	5.2	4.9	4.9
Heteroskedasticity:								
100 obs	6.8	43.7	13.2	26.9	6.2	7.5	12.0	12.0
500 obs	4.7	47.1	8.7	26.3	5.2	5.1	8.3	8.3
χ^2_{10}:								
100 obs	4.8	7.8	13.1	10.5	12.4	11.0	9.9	9.9
500 obs	6.5	5.6	54.3	42.0	67.9	58.6	22.3	22.3
t_{10}:								
100 obs	5.0	6.9	10.5	9.3	8.3	8.2	9.5	9.5
500 obs	4.4	6.6	16.7	15.3	6.2	7.7	17.2	17.2
b) Ordered Probit:								
df:	1	1	2	4	3	6	1	5
Omitted Variable:								
100 obs	50.6	5.9	6.9	34.8	5.8	5.5	4.3	5.6
500 obs	52.1	6.6	4.2	33.2	6.8	5.4	6.7	6.1
Heteroskedasticity:								
100 obs	7.0	59.7	13.4	44.0	8.3	13.3	8.9	7.9
500 obs	6.5	58.7	8.7	39.1	5.9	4.5	6.1	8.0
χ^2_{10}:								
100 obs	4.1	9.6	32.6	27.1	19.8	18.6	6.7	25.3
500 obs	6.0	11.3	91.3	86.6	76.5	58.9	15.7	81.6
t_{10}:								
100 obs	4.5	8.1	18.9	17.4	14.6	15.6	9.4	13.0
500 obs	6.8	9.7	33.9	32.4	18.4	24.9	11.7	21.8

Note: Shown are rejection percentages in 1000 replications.

TABLE VII

Powers in Logit and Ordered Logit Model
Using Estimated 5% Critical Values

Test:	LM-1	LM-2	LM-3	LM-4	LM-5	LM-6	χ^2-1	χ^2-2	IM-1	IM-2
a) Logit:										
df	1	1	2	4	1	3	0	1	1	1
Omitted Variable:										
100 obs	45.8	3.6	3.7	21.1	3.5	31.4	-	4.5	3.9	3.9
500 obs	44.0	5.8	3.8	23.8	4.8	30.2	-	4.6	7.2	7.2
Heteroskedasticity:										
100 obs	4.0	42.9	7.7	18.8	5.8	26.0	-	5.6	8.8	8.8
500 obs	4.4	48.5	5.2	23.9	5.5	30.3	-	4.4	7.0	7.0
χ^2_{10}:										
100 obs	4.3	4.6	4.8	5.0	4.3	4.4	-	5.8	3.3	3.3
500 obs	4.5	5.2	25.4	16.4	50.7	26.3	-	7.4	9.3	9.3
t_{10}:										
100 obs	4.5	5.3	2.3	2.8	2.9	3.3	-	5.7	2.4	2.4
500 obs	4.1	5.2	3.0	4.1	3.6	4.3	-	5.4	7.5	7.5
b) Ordered Logit:										
df	1	1	2	4	1	3	3	6	1	5
Omitted Variable:										
100 obs	48.8	4.1	4.2	23.6	5.2	33.3	4.5	3.5	5.0	3.5
500 obs	57.1	4.5	4.1	25.1	4.3	34.3	3.7	4.5	4.0	4.9
Heteroskedasticity:										
100 obs	5.1	59.0	10.1	38.0	7.7	47.1	2.7	1.8	4.6	4.2
500 obs	5.4	54.3	5.9	29.1	4.8	40.0	3.8	5.1	4.7	7.0
χ^2_{10}:										
100 obs	5.6	5.1	12.4	11.2	21.0	15.4	10.3	9.2	3.4	11.4
500 obs	5.6	4.9	54.9	39.8	88.6	73.5	27.9	26.2	5.1	60.6
t_{10}:										
100 obs	5.9	4.6	3.9	3.9	3.0	5.6	3.8	4.0	5.1	2.1
500 obs	6.4	4.0	3.5	4.3	3.5	5.4	3.9	4.2	5.3	4.8

Note: Shown are rejection percentages in 1000 replications.

grees of freedom. (The means of LM-3 were larger than those of LM-5). At the same time, the extra term in LM-3 does not give power against the extra kurtosis in the t alternative. The chi-squared tests only seem to have power against the alternative of asymmetric errors. IM-1 is also poor, and IM-2 evidently has power only against non-logistic and non-normal errors. This behavior of the IM tests might be expected given that the generalized residuals do not enter into the test directly and the design of the experiments means that the heteroskedasticity weights in (30), the x_{1i}^2, are quite different from the ν_i. The powers of the chi-squared and IM tests are lower in the OL than the OP models, perhaps because of the logistic distribution has fatter tails than the normal.

8. Concluding Comments.

I have derived the expressions for various specifications tests in the ordered logit and probit models. As might be expected, most of the results are fairly standard and extend those in Chesher and Irish (1987). The complications are those introduced by having to estimate the thresholds γ_j^*.

The Monte Carlo experiment into the small sample properties of the tests indicated that OPG forms of the tests should not be used. The LM tests were generally preferable to the other tests considered and, in fact, the other tests had little power against alternatives other than the asymmetric error distributions. Examples of the use of the tests may be found in Weiss (1992).

ACKNOWLEDGEMENTS

I thank the two referees and the participants of the 1994 Australasian Meetings of the Econometric Society for comments and Stan Panis for research assistance. Any errors are my own.

REFERENCES

Albert, A., and J.A. Anderson (1984), "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika*, 71, 1-10.

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

Andrews, D.W.K. (1988), Chi-Squared Diagnostic Tests for Econometric Models: Theory," *Econometrica*, 56, 1419-1453.

Bera, A.K., Jarque, C.M., and L-F. Lee (1984), "Testing the Normality Assumption in Limited Dependent Variable Models," *International Economic Review*, 25, 563-578.

Chesher, A. (1984), "Testing for Neglected Heterogeneity," *Econometrica*, 52, 865-872.

Chesher, A., and M. Irish (1987), "Residual Analysis in the Grouped and Censored Normal Linear Model," *Journal of Econometrics*, 34, 33-61.

Davidson, R., and J.G. MacKinnon (1983), "Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test," *Economics Letters* 12, 269-275.

Davidson, R., and J.G. MacKinnon (1984), "Convenient Specification Tests for Logit and Probit Models," *Journal of Econometrics* 25, 241-262.

Gourieroux, C., Monfort, A., Renault, E., and A. Trognon (1987), "Generalised Residuals," *Journal of Econometrics*, 34, 5-32.

Greene, W.H. (1993), *Econometric Analysis*, second edition, Macmillan, New York.

Johnson, N.L., and S. Kotz (1970), *Continuous Univariate Distributions-1*, Wiley, New York.

Lancaster, T. (1984), "The Covariance Matrix of the Information Matrix Test," *Econometrica* 52, 1051-1053.

Moon, C-G. (1988), "Simultaneous Specification Test in a Binary Logit Model: Skewness and Heteroscedasticity," *Communications in Statistics – Theory and Methods*, 17, 3361-3387.

Newey, W.K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047-1070.

Pagan, A. and F. Vella (1989), "Diagnostic Tests for Models Based on Individual Data: A Survey" *Journal of Applied Econometrics*, 4, 529-559.

Pratt, J.W. (1981), "Concavity of the Log Likelihood," *Journal of the American Statistical Association*, 76, 103-106.

Ruud, P. (1984), "Tests for Specification in Econometrics," *Econometric Reviews*, 3, 211-242.

Skeels, C.L. and F. Vella (1991), "A Monte Carlo Investigation of the Performance of Conditional Moment Tests in Tobit and Probit Models," paper presented at the Australasian Meetings of the Econometric Society.

Weiss, A.A. (1992), "The Effects of Helmet Use on the Severity of Head Injuries in Motorcycle Accidents," *Journal of the American Statistical Association*, 87, 48-56.

White, H. (1982), "Maximum Likelihood Estimation in Misspecified Models," *Econometrica*, 50, 1-25.