Journal of Econometrics 33 (1986) 341-365. North-Holland

SPECIFICATION AND TESTING OF SOME MODIFIED COUNT DATA MODELS*

John MULLAHY

Yale University, New Haven, CT 06520, USA
Resources for the Future, Washington, DC 20036, USA

Received October 1984, final version received April 1986

This paper explores the specification and testing of some modified count data models. These alternatives permit more flexible specification of the data-generating process (dgp) than do familiar count data models (e.g., the Poisson), and provide a natural means for modeling data that are over- or underdispersed by the standards of the basic models. In the cases considered, the familiar forms of the distributions result as parameter-restricted versions of the proposed modified distributions. Accordingly, score tests of the restrictions that use only the easily-computed ML estimates of the standard models are proposed. The tests proposed by Hausman (1978) and White (1982) are also considered. The tests are then applied to count data models estimated using survey microdata on beverage consumption.

1. Introduction

Interest in comer-solution problems in econometrics has given rise to an assortment of methods designed to allow consistent parameter estimation in a corresponding assortment of model specifications. Recent extensions of the econometric research on comer-solution outcomes have concentrated on the specification and testing of models for non-negative data that are measured as integers, or count data.[1] In addition to discussion of basic model structures,

*This paper is a substantial revision of a paper originally circulated as 'Hurdle Models for Discrete and Grouped Dependent Variables'. The research has been supported in part by a cooperative agreement between the U.S. Environmental Protection Agency and Resources for the Future but should not be inferred to represent views of EPA. Thanks are due to two referees and Paul Portney for helpful and constructive comments on earlier drafts. Any errors that remain must be attributed to the author.

[1] Much of the interest in count data modeling appears to stem from the recognition that the use of continuous distributions to model integer outcomes might have unwelcome consequences, including inconsistentparameter estimates. Even if it is maintained that the integer outcomes are generated by latent continuous variates, the arguments presented by Stapleton and Young (1984) suggest that because such integer outcomes are actually continuous realizations measured with error, inconsistent parameter estimates can result if standard methods like ML Tobit are used for estimation. Rosenzweig and Wolpin (1982) explicitly rule out the use of a Tobit estimator in their analysis of fertility outcomes, in which the fertility measure used is the number of children born to a mother during some time interval. In another application, Portney and Mullahy (1986) conduct some tests for a Tobit specification of their count measure and find considerable evidence of misspecification.

the interesting complications introduced by random and fixed effects, overdispersion, distributional misspecification, censoring, and multivariate dependent variables have also been treated in the econometrics literature.[2]

One issue yet to be addressed in detail is whether the statistical model governing the binary outcome of the count being either zero or positive might differ from that determining the magnitude of the positive counts. In standard[3] count data models familiar to economists (e.g., the Poisson), these two processes are constrained to be identical. That is, letting $\phi_1(y,\theta_1)$ and $\phi_2(y,\theta_2)$ be two functions defined on $y \in \Gamma = \{0,1,2\ldots\}$ satisfying $\phi_1, \phi_2 > 0$ and

$$\phi_1(0,\theta_1) + \sum_{y \in \Gamma_+} \phi_2(y,\theta_2) = 1, \tag{1}$$

where $\Gamma_+ = \Gamma \setminus \{0\}$, a standard count data model specifies $\phi_1(y,\theta_1) = \phi_2(y,\theta_2)$ for all $y \in \Gamma$, so that

$$\sum_{y \in \Gamma} \phi_1(y,\theta_1) = \sum_{y \in \Gamma} \phi_2(y,\theta_2) = 1. \tag{2}$$

Proposed here are two types of modifications to the basic count data models in which (1) is satisfied but where $\phi_1(y,\theta_1) \neq \phi_2(y,\theta_2)$. These are termed hurdle[4] and with-zeros (WZ)[5] models. While the two types of modifications in general have different structures, it is shown later that they collapse into the same model under some circumstances. The basic idea underlying these modifications is that both permit the relative probabilities of zero and non-zero realizations to differ from those implied by the parent distributions that they modify.

A particularly interesting feature of the modified count data specifications considered here is that they provide a natural means for modeling overdispersion or underdispersion of the data.[6] Specifically, overdispersion and underdispersion are viewed as arising from a misspecification of the maintained parent dgp in which the relative probabilities of zero and non-zero (positive) realizations implied by the parent distribution are not supported by the data. By

[2] See variously Cameron and Trivedi (1986), Gourieroux, Montfort and Trognon (1984b), Hausman, Hall and Griliches (1984), Hausman, Ostro and Wise (1984), Lee (1984a), Manning, Lillard and Phelps (1983), and Terra (1985).

[3] The adjectives 'standard, 'basic', and 'parent' when describing count data models are used interchangeably in this paper, and refer to models having structures like that in eq. (2) below.

[4] This term is borrowed from Cragg (1971).

[5] This term is used by Johnson and Kotz (1969, p. 205).

[6] Cox (1983) and McCullagh and Nelder (1983) are good references on overdispersion. Section 2 treats the issue in greater detail, and provides additional citations.

permitting flexible specification of the relative probabilities of zeros and positives, the modified distributions represent alternatives to standard methods for modeling overdispersion and underdispersion' that might be appealing in a variety of circumstances.

The paper is organized as follows. Section 2 describes the basic models and the proposed alternatives. The ideas are illustrated in the context of the Poisson and geometric distributions; however, the results can be extended in a straightforward manner to other count data models of interest. Section 3 discusses alternative strategies for specification testing. Section 4 illustrates the ideas with count data models estimated using survey microdata on beverage consumption. Section 5 summarizes the paper.

2. Some modified count data models

This section considers the specification of two types of modified count data models. First, the basic Poisson and geometric models, which are used to focus the analysis, are summarized. The two types of modifications are then described in detail. Finally, the relationships between overdispersion and the modified distributions are discussed.

The Poisson distribution of count variate Y_t is defined by

$$\begin{aligned} P(y,\lambda) &= \exp(-\lambda)\lambda^y/y!, \quad && y \in \Gamma, \\ &= 0, && \text{else,} \end{aligned} \tag{3}$$

with

$$\lambda = \mathrm{E}(Y_t) = \mathrm{var}(Y_t) > 0.$$

Since $\lambda > 0$ the influence of covariates is admitted by specifying $\lambda_t = \exp(X_t\beta)$, X_t' and β here and throughout being $k \times 1$ vectors of fixed covariates and unknown parameters, respectively. Thus, the loglikelihood function for a sample of T independent observations (suppressing terms not depending on β) is[8]

$$L^{\mathrm{P}} = \sum_{t\in\Omega} y_t X_t\beta - \exp(X_t\beta), \tag{4}$$

'For example, a popular generalization of the Poisson model assumes that a random component in the basic Poisson expectation function is gamma-distributed in the population, so that a negative binomial model results as a gamma mixture of the Poissons. See Hausman, Hall and Griliches (1984) and Cameron and Trivedi (1986) for additional discussion.

[8] In the following, $f_\delta = \partial f/\partial\delta$ denotes the vector of first partial derivatives of j with respect of δ, and $f_{\delta\zeta} = \partial^2 f/\partial\delta\partial\zeta'$ is the matrix of second partials of f with respect to δ and ζ'. Where no ambiguity is possible, ∇f and $\nabla^2 f$ are occasionally used to denote the gradient and Hessian of f. $\Omega_0 = \{t|y_t = 0\}$, $\Omega_1 = \{t|y_t \in \Gamma_+\}$, and $\Omega = \Omega_0 \cup \Omega_1$. The symbols L and A refer in general to loglikelihood functions of basic and modified models, respectively.

where y_t is the count for the t th observation. The ML estimate of β satisfies

$$L_\beta^{\mathrm{P}} = \sum_{t \in \Omega} [y_t - \exp(X_t\beta)] X_t' = 0, \tag{5}$$

and the Hessian of L^{P} is

$$L_{\beta\beta}^{\mathrm{P}} = \sum_{t \in \Omega} - \exp(X_t\beta) \mathrm{X/X,}. \tag{6}$$

Should circumstances suggest an alternative to the Poisson distribution be considered, the geometric distribution is one possible candidate.' The geometric model is not encumbered by the Poisson's mean = variance property, as the variance of a geometric variate exceeds its mean. The geometric characterizes discrete decay phenomena in the sense that its probabilities obey Pr(y) > Pr(y + 1) for all y $\in \Gamma$. Thus, the Poisson distribution is in one sense both more flexible than the geometric, since non-decay count models [i.e., Pr(y) < Pr(y + 1) for some y] can be admitted, and more restrictive, since such decay processes obtain only for $\lambda < 1$ in the Poisson.

The geometric distribution of a count variate Y_t is defined by

$$\begin{aligned} G(y, \gamma) &= \gamma^y (1+\gamma)^{-(y+1)}, \quad y \in \Gamma \\ &= 0, \qquad\qquad\qquad\qquad \text{else,} \end{aligned} \tag{7}$$

where

$$\mathrm{E}(Y_t) = \gamma \quad \text{and} \quad \operatorname{var}(Y_t) = \gamma(1+\gamma).$$

Since $\gamma > 0$, $\gamma_t = \exp(X_t\beta)$ is the obvious parameterization. The loglikelihood function can thus be written as

$$L^{\mathrm{G}} = \sum_{t \in \Omega} y_t X_t \beta - (y_t + 1)\log[1 + \exp(X_t\beta)]. \tag{8}$$

The ML estimate of β satisfies

$$L_\beta^{\mathrm{G}} = \sum_{t \in \Omega} \{[y_t - \exp(X_t\beta)]/[1 + \exp(X_t\beta)]\} X_t' = 0, \tag{9}$$

and the Hessian is

$$L_{\beta\beta}^{\mathrm{G}} = \sum_{t \in \Omega} - \{(y_t + 1)\exp(X_t\beta)/[1 + \exp(X_t\beta)]^2\} X_t' X_t. \tag{10}$$

'Discussion is confined here to the geometric version of the negative binomial. The analysis is extended to the general negative binomial distribution and other count data models in a straightforward manner.

The first modified count data models considered here are termed hurdle models, following the terminology developed by Cragg (1971).[10] The idea underlying the hurdle formulations is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is positive, the 'hurdle' is crossed, and the conditional distribution of the positives is governed by a truncated-at-zero count data model. Formally, for a random variable Y, the conditional distribution of the positives is $\phi_2(y,\theta_2)/\Phi_2(\theta_2), y \in \Gamma_+$, where ϕ_2 satisfies (2) and Φ_2, the summation of ϕ_2 on the support of the conditional density, is the truncation normalization. The probability that the threshold is crossed is $\Phi_1(\theta_1)$.[11] Thus, the general form of the hurdle model likelihood function is

$$\exp(\Lambda^{\mathrm{H}}) = \prod_{t \in \Omega_0} [1 - \Phi_1(\theta_1)] \prod_{t \in \Omega_1} [\phi_2(y,\theta_2)\Phi_1(\theta_1)/\Phi_2(\theta_2)]. \tag{11}$$

When $\Phi_1(\theta_1) = \Phi_2(\theta_2)$, (11) reduces to

$$\exp(\Lambda^{\mathrm{H}}) = \prod_{t \in \Omega_0} [1 - \Phi_2(\theta_2)] \prod_{t \in \Omega_1} [\phi_2(y,\theta_2)], \tag{12}$$

which resembles the likelihood function of a Tobit model. If $\Phi_1(\theta_1) = \Phi_2(\theta_2)$ as a result of parameter restrictions $\theta_1 = \theta_2$, then the model is akin to that investigated for normal distributions by Cragg and by Lin and Schmidt (1984), who demonstrate that Tobit results as a parameter-restricted version of one of Cragg's original specifications.

In any particular application, there will likely exist numerous plausible specifications of both the binary probability model and the conditional distribution of the positives. For present purposes, only specifications where (11) reduces to (12) as a result of the parameter restrictions $\theta_1 = \theta_2$ are of interest, the objective here being the development of count data analogs of Cragg's Tobit modifications.

To motivate the Poisson hurdle specification, consider the dgp:

$$\Pr(y=0) = \exp(-\lambda_1)\lambda_1^y/y! = \exp(-\lambda_1), \tag{13}$$

$$[1 - \Pr(y=0)] = \sum_{y \in \Gamma_+} \Pr(y) = [1 - \exp(-\lambda_1)], \tag{14}$$

and

$$\begin{aligned} \Pr(y|y>0) &= \lambda_2^y/\{[\exp(\lambda_2)-1]y!\}, \quad && y \in \Gamma_+, \\ &= 0, && \text{else}, \end{aligned} \tag{15}$$

[10] See Cragg *(1971)* or Lin and Schmidt (1984) for additional discussion.

[11] In general, Φ_1, ϕ_2, and Φ_2 also depend on covariates X_t.

where λ_1 is the parameter of a Poisson/exponential distribution governing the probability of observing a positive count and where (15) has the form of a truncated- or positive-Poisson distribution. Parameterizing λ_{jt} as $\exp(X_t\beta_j)$, the loglikelihood function based on (13)–(15) is

$$\Lambda^{\mathrm{PH}} = \log\Bigg\{\Bigg[\prod_{t\in\Omega_0}\{\exp[-\exp(X_t\beta_1)]\}\prod_{t\in\Omega_1}\{1-\exp[-\exp(X_t\beta_1)]\}\Bigg]$$

$$\times\Bigg[\prod_{t\in\Omega_1}\exp(y_tX_t\beta_2)/(\{\exp[\exp(X_t\beta_2)]-1\}y_t!)\Bigg]\Bigg\} \qquad (16)$$

$$= [\Lambda^{\mathrm{P1}}(\beta_1)] + [\Lambda^{\mathrm{P2}}(\beta_2)],$$

which reduces to (4) when $\beta_1 = \beta_2$. Λ^{P1} can be regarded as a loglikelihood function for the binary (zero/positive) outcome and Λ^{P2} as a loglikelihood function for a truncated-Poisson model. Thus, the ML estimates of β_1 and β_2 can be obtained by separate maximization of Λ^{P1} and Λ^{P2}, respectively.

For the geometric model, an interesting hurdle specification corresponding to (11) is the dgp:

$$\Pr(y=0) = 1/(1+\gamma_1), \qquad (17)$$

$$[1-\Pr(y=0)] = \sum_{y\in\Gamma_+}\Pr(y) = \gamma_1/(1+\gamma_1), \qquad (18)$$

and

$$\Pr(y|y>0) = \gamma_2^{(y-1)}/[(1+\gamma_2)^y], \quad y\in\Gamma_+$$
$$= 0. \qquad \text{else.} \qquad (19)$$

Parameterizing y,, = $\exp(X_t\beta_j)$, it is seen that the binomial probabilities (17) and (18) are identically those of a standard binomial logit model.[12] Eq. (19) is in the form of a truncated-at-zero geometric model. The complete loglikeli-

[12] This result suggests that when the basic specification is correct, the geometric parameters can be estimated consistently, though not efficiently, by standard binomial logit programs. Such an approach is analogous to consistent but inefficient estimation of a Tobit model's parameters using a probit model [see Amemiya (1984) and Ruud (1984)]. In the Tobit/probit case, however, only the scaled parameters β/σ can be estimated by probit. Due to the functional dependence of the location and scale parameters in the geometric specification, the natural or unscaled parameters can be estimated by logit. Appendix A demonstrates the relative inefficiency of both the logit and truncated geometric estimators of the basic geometric model.

(17)–(19) is

$$\begin{aligned}\Lambda^{GH} = \log\Bigg\{ & \Bigg[\prod_{t\in\Omega_0} \{1/[1+\exp(X_t\beta_1)]\} \\ & \times \prod_{t\in\Omega_1} \{\exp(X_t\beta_1)/[1+\exp(X_t\beta_1)]\} \Bigg] \\ & \times \Bigg[\prod_{t\in\Omega_1} \exp[(y_t-1)X_t\beta_2]/\{[1+\exp(X_t\beta_2)]^{y_t}\} \Bigg] \Bigg\} \\ = {} & [\Lambda^{G1}(\beta_1)] + [\Lambda^{G2}(\beta_2)],\end{aligned} \tag{20}$$

which reduces to (8) when $\beta_1 = \beta_2$. Again the ML estimates can be obtained by separate maximization of Λ^{G1} and Λ^{G2}.

The second class of modified count models is termed the WZ class, following the terminology developed by Johnson and Kotz (1969, pp. 204-206). Like the hurdle models, the idea motivating the WZ specifications is that the conditional distribution of the positives is properly characterized by the truncated-at-zero version of the parent distribution. The probabilities of the positives relative to the probability of the zero outcome, however, are no longer as specified by the parent distribution. Instead, the WZ model specifies that the probability of the zero outcome is additively augmented or reduced by an amount ψ so that, in the notation of (1) and (2),

$$\phi_1(y,\psi,\theta) = \psi + (1-\psi)\phi(y,\theta), \qquad y = 0, \tag{21}$$

and

$$\phi_2(y,\psi,\theta) = (1-\psi)\phi(y,\theta), \qquad y \in \Gamma_+, \tag{22}$$

where $\phi(y,\theta)$ are the probabilities specified for all $y \in \Gamma$ by the parent density, and the terms $(1-\psi)$ ensure that (21) and (22) constitute a proper discrete probability distribution.[13] When $\psi > 0$ ($\psi < 0$), the relative probabilities $\phi_1(0,\theta)/\phi_2(y,\theta)$, $y \in \Gamma_+$, are greater (less) than those specified by the parent distribution; similarly, when $\psi > 0$ ($\psi < 0$) $E(Y)$ is less (greater) than in the parent model. When $\psi = 0$, the basic distribution obtains.

The loglikelihood functions of the Poisson and geometric WZ models are, respectively,

$$\begin{aligned}\Lambda^{PZ} = & \sum_{t\in\Omega_0} \log\{\psi + (1-\psi)\exp[-\exp(X_t\beta)]\} \\ & + \sum_{t\in\Omega_1} \log(1-\psi) - \exp(X_t\beta) + y_t X_t \beta,\end{aligned} \tag{23}$$

[13] Johnson and Kotz *(1969,* p. *205)* note that the constraint $\psi \in \Psi = [-\phi_0/(1-\phi_0), 1)$, where $\phi_0 \equiv \phi(0,\theta)$, is also required.

and

$$\Lambda^{GZ} = \sum_{t\in\Omega_0} \log[\psi \exp(X_t\beta) + 1]$$

$$-\log[1 + \exp(X_t\beta)] + \sum_{t\in\Omega_1} \log(1-\psi) \tag{24}$$

$$+ y_t X_t\beta - (1+y_t)\log[1+\exp(X_t\beta)],$$

where terms not depending on (β,ψ) are suppressed.[14]

Although the hurdle and WZ specifications represent different modifications of the basic count data models, it is interesting that they collapse into the same specification in the case where only an intercept is included in the X_t vectors. Here the basic models specify E(Y) = $\exp(\beta)$ and the hurdle models have parameters $\beta_1 = \beta + \alpha$ and $\beta_2 = \beta$, where β, β_1, β_2, and α are scalars. Some manipulation of (16) and (23) yields

$$\psi = \{1 - \exp[\exp(\beta) - \exp(\beta+\alpha)]\}/\{1 - \exp[\exp(\beta)]\}$$

in the Poisson hurdle model, while similar manipulation of (20) and (24) gives

$$\psi = [1 - \exp(\alpha)]/[1 + \exp(\beta+\alpha)]$$

in the geometric hurdle model. In both instances, sign(a) = $-\text{sign}(\psi)$, and, for finite β, ψ approaches $\sup(\Psi)$ and $\inf(\Psi)$ as α approaches $-$ cc and $+\infty$, respectively. It is also the case in the intercept-only specifications that the ML estimate of the intercept parameter in the WZ models is identical to that obtained by ML estimation of the intercept parameter in the truncated-at-zero variant of the parent distribution.[15]

Overdispersion in count data models has been discussed extensively.[16] Overdispersion is meaningful only in reference to some maintained dgp for y, and for present purposes can be defined as a situation where the ratio $\text{var}(Y)/\text{E}(Y)$ exceeds that implied by the maintained dgp for y. For example, overdispersion is present in the basic Poisson and geometric models if $\text{var}(Y)/\text{E}(Y) > 1$ and $\text{var}(Y)/\text{E}(Y) > 1 + \text{E}(Y)$, respectively. Underdispersion is defined by reversing the inequalities.

To see that the hurdle models naturally admit overdispersion or underdispersion, consider $\text{var}(Y)/\text{E}(Y)$ in a general hurdle formulation. In the nota-

[14] The gradients and Hessians of (20) and (24) are presented in appendix B.

“Johnson and Kotz (1969, pp. 205-206).

[16] See, for example, Cox (1983), Hausman, Hall and Griliches (1984), Cameron and Trivedi (1986), and McCullagh and Nelder (1983).

tion of (11), if the dgp of y has the hurdle structure then

$$\mathrm{var}(Y)/\mathrm{E}(Y) = \left\{ \sum y^2\phi_2(y,\theta_2) - (\Phi_1/\Phi_2)\left[\sum y\phi_2(y,\theta_2)\right]^2\right\} \Big/ \left\{ \sum y\phi_2(y,\theta_2)\right\}, \tag{25}$$

where each summation is over $y \in \Gamma$. If $(\Phi_1/\Phi_2) = 1$, then var(Y)/E(Y) is identical to that given by the basic model. When $(\Phi_1/\Phi_2) \neq 1$, the hurdle formulation characterizes overdispersion when $(\Phi_1/\Phi_2) \in (0,1)$ and underdispersion when $(\Phi_1/\Phi_2) \in (1, +\infty)$.

The WZ specifications allow for overdispersion and underdispersion in a similar manner, viz,

$$\mathrm{var}(Y)/\mathrm{E}(Y) = \left\{\sum y^2\phi(y,\theta) - (1-\psi)\left[\sum y\phi(y,\theta)\right]^2\right\} / \left\{\sum y\phi(y,\theta)\right\}, \tag{26}$$

where the summation is again over $y \in \Gamma$, and the $\phi(y,\theta)$ are as specified in the basic model. Overdispersion and underdispersion are present as $\psi \in (0,1)$ and $\psi \in [-\phi_0/(1-\phi_0), 0)$, respectively, while the basic models result when $\psi = 0$.

Thus, tests of the null hypotheses $\theta_1 = \theta_2$ (i.e., $\Phi_1 = \Phi_2$) in the hurdle models or $\psi = 0$ in the WZ models are implicitly tests of the null hypothesis of no overdispersion or underdispersion of the kind described here. It should be noted, however, that overdispersion and underdispersion can be manifested in forms other than those examined here.

3. Specification testing

This section discusses several specification tests for the models described in section 2. Score tests and Hausman (1978) tests are proposed for testing the basic specifications against specific alternatives, the hurdle and WZ models. White's (1982) information matrix test is proposed as an omnibus test of the null hypothesis that the basic specification is a correct characterization of the dgp.

The score (or Lagrange multiplier) principle for specification testing in econometrics has been discussed extensively.[17] Because ML estimates of the

[17] See Breusch and Pagan (1980) and Engle (1984) for detailed discussions.

basic count data models can be easily obtained,[18] the score test approach is appealing here. The general form of the score statistic for testing H_0: $h(\theta)=0$ is

$$\xi = s(\hat{\theta})'\Upsilon(\hat{\theta})^{-1}s(\hat{\theta}), \tag{27}$$

where $s(\hat{\theta})$ is the $k \times 1$ score vector and $\Upsilon(\hat{\theta})$ is the $k \times k$ information matrix, both evaluated at the ML estimates of the restricted model; $\theta = (\theta_1', \theta_2')'$ is the $k \times 1$ parameter vector, where θ_1 is $p \times 1$ and θ_2 is $(k-p) \times 1$; and $h(\theta)$ is an $r \times 1$ vector of restrictions, where for present purposes $h(\theta) = \theta_1 = 0$ and $p = r$.[19] Since some elements of $s(\hat{\theta})$ are identically zero, only the non-zero subvector of $s(\hat{\theta})$ and corresponding submatrix of $\Upsilon(\hat{\theta})^{-1}$ are required to compute ξ. Under H_0, ξ is asymptotically distributed as a central χ^2_p variate.

Much discussion of the score test has focused on computational methods.[20] Given $s(\hat{\theta})$ and $\Upsilon(\hat{\theta})$, ξ can of course be computed using matrix calculations according to (27). Typically Υ will be estimated by either the negative Hessian of the restricted loglikelihood function or by the gradient outer product $G'G$, where G is the $T \times k$ matrix having typical element $[\partial \Lambda_t / \partial \theta_j]$, evaluated in either case at the restricted ML estimates. Alternatively, ξ can often be obtained as a function of the R^2 of some auxiliary linear regression.[21]

Computations of ξ based on different estimates of Υ will yield different values of the test statistic in finite samples, even when the null hypothesis is true and the model is correctly specified. A separate complication arises when the probabilities underlying the model's likelihood function are misspecified. Although consistent parameter estimation under such circumstances is possible when the expectation function has been correctly specified,[22] inferences based on standard estimates of the parameter covariance matrix will generally not be robust against such misspecification. White (1982) and Engle (1984) have suggested an amendment to the standard form of the score statistic (27)

[18] Since the loglikelihood functions of both the basic Poisson and basic geometric models (4) and (8) are concave, convergence to the ML estimates using a Newton–Raphson algorithm has proven in practice to be quite rapid. Alternatively, non-linear weighted least squares can be used to obtain the ML estimates of these models; see Hausman, Hall and Griliches (1984) and Hausman, Ostro and Wise (1984).

[19] Since the paper is ultimately concerned with applying the finite-sample analogs of these test statistics, whose known properties are largely asymptotic, Υ is taken here to be $-(\nabla^2\Lambda)$ rather than $-\mathrm{E}(\nabla^2\Lambda/T)$.

[20] See Engle (1984) and Davidson and MacKinnon (1984a, b).

[21] For example, since Υ can be estimated by $G'G$, ξ can be calculated as

$$\xi = \iota' G(\hat{\theta})(G(\hat{\theta})'G(\hat{\theta}))^{-1}G(\hat{\theta})'\iota \tag{*}$$

where ι is a $T \times 1$ vector of ones. Since $\iota' G(\hat{\theta}) = s(\hat{\theta})'$, (*) is simply an alternative expression of (27). Moreover, since $\iota'\iota = T$, ξ in (*) is seen to be T times the uncentered R^2 from the regression of ι on $G(\hat{\theta})$, or, alternatively, $\tilde{\iota}'\tilde{\iota}$ from the same regression.

[22] See Gourieroux, Montfort and Trognon (1984a) and Cameron and Trivedi (1986).

that ensures a test of the proper size when such misspecification is present. Defining $A = -\nabla^2\Lambda(\hat{\theta})$ where Λ is the maintained loglikelihood function, $B = G(\theta)'G(\hat{\theta})$, and $C = A^{-1}BA^{-1}$, then the finite-sample analog of the statistic proposed by White and Engle is

$$\xi^* = s_1(\hat{\theta})'A^{11}(C_{11})^{-1}A^{11}s_1(\hat{\theta}), \tag{28}$$

where the $(1,1)$ blocks of A^{-1} and C correspond to the p non-zero elements of $s(\hat{\theta})$, $s_1(\hat{\theta}) = [\partial\Lambda/\partial\theta_1]|_{(\theta=\hat{\theta})}$. For purposes of comparison, the empirical illustrations presented below in section 4 present score statistics calculated using both standard parameter covariance estimates [i.e., $(-\nabla^2\Lambda)^{-1}$ and $(G'G)^{-1}$] and the approach suggested by White and Engle.

The score test strategy for the count data hurdle models draws conceptually on the work of Lin and Schmidt (1984). It was demonstrated in section 2 that the basic count model specifications result when the restriction $\beta_1 = \beta_2$ is imposed in the hurdle models. Computation of the score test statistic is simplified by reparameterizing the hurdle models along the lines suggested by Lin and Schmidt where given the new parameters (α, β), with $\beta_1 = \alpha + \beta$ and $\beta_2 = \beta$, the score test is of H_0: $\alpha = 0$. Under this reparameterization, however, the Hessian of the hurdle model loglikelihood function is no longer block-diagonal (see appendix B).

The score test for the WZ specifications is of H_0: $\psi = 0$. Since H_0 specifies a point in $\text{int}(\Psi)$, standard methods of inference can be used. Note that only one element of $s(\hat{\theta})$ is non-zero since ψ is scalar. Under H_0, ξ is asymptotically distributed as χ_1^2. The computation of the score test is complicated, however, because the Hessian of the WZ loglikelihood function is not block-diagonal. Appendix B provides the formulae used to compute the geometric hurdle and WZ model score tests.

An alternative test strategy for the hurdle models recognizes that when $\beta_1 = \beta_2$ in (16) or (20), β $(= \beta_1 = \beta_2)$ can be estimated consistently by maximizing the full loglikelihood function Λ^{jH} (j = P, G), or either of its components $(\Lambda^{j1}, \Lambda^{j2})$. However, as noted earlier and demonstrated in appendix A for the geometric model, the latter estimates are inefficient relative to the former. Of course, when $\beta_1 \neq \beta_2$, the three estimators will diverge asymptotically. These properties suggest that a Hausman test approach can be used to test H_0: $\beta_1 = \beta_2$.

A finite-sample version of the Hausman test statistic is used here:

$$H = (\hat{\beta}_u - \hat{\beta})'(\hat{V}(\hat{\beta}_u) - \hat{V}(\hat{\beta}))^{-1}(\hat{\beta}_u - \hat{\beta}), \tag{29}$$

where $\hat{\beta}$ is the ML estimate of the restricted (i.e., basic) model, $\hat{\beta}_u$ is either of the two estimates of the parameters of the unrestricted model (β_1 or β_2), and

$\hat{V}(\hat{\beta})$ and $\hat{V}(\hat{\beta}_u)$ are estimates of the corresponding covariance matrixes. Under $\mathbf{H}_0$: $\beta_1 = \beta_2$, the asymptotic analog of H is distributed $\chi^2_{(.5k)}$. This test is particularly appealing in the case of the geometric distribution, where one estimate of $(\beta_u, V(\hat{\beta}_u))$ is easily obtained using familiar logit techniques.

White (1982) has analyzed estimation in cases where the wrong probability model is used to construct what the researcher believes to be the likelihood function. Such misspecification is destined to plague many applications of economic count data models, as theory will typically suggest little about the model's probability structure.[23] Accordingly, the information matrix (IM) test developed by White is a potentially valuable diagnostic tool in the analysis of economic count data models. Although both the IM and score tests use only the ML estimates of the parameters of the basic model, the IM test differs from the score tests proposed above since no specific alternative specifications or parameter restrictions are involved in its computation. The IM test principle relies solely on properties of the specification being tested that must obtain under the null hypothesis of no misspecification. As such, the IM test is an omnibus specification test; rejection of the null hypothesis of no misspecification would not specifically favor either the hurdle or the WZ variant, or any other specific alternative for that matter.[24]

4. Empirical analysis

To illustrate the specifications and tests described above, data from the 1980 Wave II of the National Survey of Personal Health Practices and Consequences (NSPHPC) are used.[25] Among the data reported in the NSPHPC are individuals' daily consumption of various beverages. Although beverage quantity is a continuous measure, the protocol in the NSPHPC is to report consumption in integer amounts (number of cups, glasses, etc.). Such beverage consumption measures serve well to illustrate the points discussed above. Analyzed here are individuals' daily consumption of coffee *(COFFEE)*, tea *(TEA)*, and milk (MILK). The explanatory variables used are an intercept (INT), age in years *(AGE)*, *years* of completed schooling *(EDUC)*, family income *(INCOME)*, and O-l dummies for sex *(MALE* = 1 if male), race (WHITE = 1 if white, = 0 if black), and marital status *(MARRIED* = 1 if

[23] Cameron and Trivedi *(1986,* p. *30).*

[24] Lee (1984a, b) notes circumstances under which the IM test is inconsistent against alternative specifications.

[25] The NSPHPC is a national, random-digit telephone survey conducted in two waves in Spring 1979 and Spring 1980. The total sample is comprised of non-institutionalized adults aged 20-64 residing in the coterminous U.S. There were 3,025 survey respondents in Wave I, 2,436 of whom also responded in Wave II. Additional details on the NSPHPC are available in U.S. Department of Commerce (1982).

Table 1

Sample frequency distribution of dependent variables ($T = 1{,}900$).

n	COFFEE	TEA	MILK
0	499	1171	767
1	259	310	557
2	341	214	333
3	235	99	142
4	189	54	62
5	123	21	23
6	97	24	16
7	12	2	0
8	53	5	0
9	3	0	0
10	58	0	0
12	13	0	0
15	18	0	0

currently married).[26] After screening for outliers,[27] the number of observations having all data necessary for estimation is 1,900. Two sets of models are estimated: the first uses the entire set of regressors, the second uses only the intercept. The sample frequency distributions of the dependent variables are presented in table 1, and the descriptive statistics for the estimation sample are presented in table 2.

For parsimony, only the geometric model and its variants are estimated here.[28] Tables 3 through 5 present the estimation results with covariates included for the coffee, tea, and milk models, respectively. The first column in each table gives the ML estimates of the basic geometric models. The second and third columns present the estimates of the unrestricted hurdle models which, as discussed earlier, can be obtained by separate ML estimation of a binary logit model estimated over the entire sample and a truncated-geometric model estimated on the sample having positive realizations of the dependent variable. The fourth column presents the ML estimates of the unrestricted geometric WZ models.[29]

[26] The education and income variables are pseudo-continuous, constructed using interval midpoints. For the open-ended intervals, the value 17 was used for the schooling category '16 or more' years, and the value 35,000 was used for the income category '$25,000 or more'. Some variables required to properly interpret the estimated models as demand functions are not available (e.g., own and substitute goods' prices); similarly, information about other determinants of beverage consumption (e.g., religion) is not provided in the NSPHPC.

[27] For example, 13 observations for which daily coffee consumption was reported as greater than 15 cups, 19 observations for which daily tea consumption was reported as greater than 8 cups, and 10 observations for which daily milk consumption was reported as greater than 6 glasses were deleted.

[28] Estimation is performed using a program written in SAS's PROC MATRIX, which is available from the author on request.

[29] For all the WZ models, the requirement that the estimate $\hat{\psi}$ be in the interval $\hat{\Psi} = [-\phi(0,\hat{\theta})/(1-\phi(0,\hat{\theta})),1)$ was found to hold for each observation in the sample.

Table 2

Sample descriptive statistics (T = 1,900)

Variable	Mean	S.D.	Min	Max
COFFEE	2.705	2.845	0.00	15.00
TEA	0.818	1.349	0.00	8.00
MILK	1.109	1.251	0.00	6.00
AGE	39.976	12.603	21.00	65.00
EDUC	13.001	2.697	0.00	17.00
INCOME	21,172.368	10,724.967	2,500.00	35,000.00
MALE	0.395	0.489	0.00	1.00
WHITE	0.931	0.253	0.00	1.00
MARRIED	0.705	0.456	0.00	1.00

Table 3

Estimation results: Dependent variable *COFFEE* (covariates included).[a]

Variable	Restricted model	Binary logit	Truncated geometric	Geometric with zeros
INT	− 0.566	− 1.419	− 0.284	− 0.676
	(0.213)	(0.390)	(0.271)	(0.205)
	[0.178]			
AGE	0.0157	0.0443	0.0061	0.0169
	(0.0023)	(0.0048)	(0.0028)	(0.0022)
	[0.0019]			
EDUC	− 0.0299	− 0.0559	− 0.0229	− 0.0308
	(0.0111)	(0.0226)	(0.0129)	(0.0106)
	[0.0092]			
INCOME	6.1E-6	1.5E-5	2.8E-6	6.5E-6
	(2.9E-6)	(5.8E-6)	(3.4E-6)	(2.8E-6)
	[2.4E-6]			
MALE	0.174	0.180	0.189	0.173
	(0.056)	(0.113)	(0.065)	(0.053)
	[0.048]			
WHITE	1.111	1.155	1.179	1.108
	(0.134)	(0.198)	(0.191)	(0.129)
	[0.120]			
MARRIED	0.041	0.062	0.021	0.045
	(0.065)	(0.125)	(0.077)	(0.062)
	[0.060]			
ψ		—	—	− 0.0701
				(0.0168)
L, Λ	− 4029.55	− 1013.17	− 2982.11	− 4021.00

[a]Figures in parentheses are estimated asymptotic standard errors derived from the negative inverse Hessian of L evaluated at the ML estimates. Figures in square brackets are estimated asymptotic standard errors derived from the parameter covariance estimates obtained using the method proposed by White (1982) and Royall (1984).

Table 4

Estimation results: Dependent variable *TEA* (covariates included).[a]

Variable	Restricted model	Binary logit	Truncated geometric	Geometric with zeros
INT	0.0018	− 0.353	0.308	0.321
	(0.2580)	(0.348)	(0.396)	(0.295)
	[0.2818]			
AGE	− 0.0046	− 0.0053	− 0.0023	− 0.0042
	(0.0028)	(0.0039)	(0.0042)	(0.0032)
	[0.0031]			
EDUC	− 0.0337	− 0.0141	− 0.0480	− 0.0371
	(0.0144)	(0.0194)	(0.0222)	(0.0166)
	[0.0154]			
INCOME	−5.1E-6	−3.1E-6	−6.5E-6	− 4.836
	(3.6E-6)	(5.0E-6)	(5.5E-6)	(4.2E-6)
	[3.9E-6]			
MALE	− 0.177	− 0.320	0.046	-0.152
	(0.072)	(0.098)	(0.108)	(0.082)
	[0.079]			
WHITE	0.457	0.292	0.602	0.470
	(0.155)	(0.198)	(0.252)	(0.173)
	[0.149]			
MARRIED	0.218	0.269	0.085	0.196
	(0.082)	(0.113)	(0.122)	(0.092)
	[0.093]			
ψ				0.263
				(0.028)
L, Λ	− 2360.79	− 1254.91	− 1067.33	− 2327.72

[a] Figures in parentheses are estimated asymptotic standard errors derived from the negative inverse Hessian of L evaluated at the ML estimates. Figures in square brackets are estimated asymptotic standard errors derived from the parameter covariance estimates obtained using the method proposed by White (1982) and Royal1 (1984).

The estimates appear generally to be plausible, and in many cases the parameters are estimated with fair precision. It is noteworthy that in almost all instances the basic model parameter estimates are bounded by the logit and truncated-geometric estimates. Also noteworthy is that, with the exception of the intercept parameter, the estimates of the basic and the WZ model parameters are quite comparable.

When interpreting the estimates of ψ as diagnostic tests for overdispersion or underdispersion, it is interesting that the results suggest the presence of overdispersion in the tea models and underdispersion in both the coffee and milk models. Since underdispersion of this nature requires the variance/mean ratio to be less than that implied by the parent model, it is possible that an

Table 5
Estimation results: Dependent variable *MILK* (covariates included).[a]

Variable	Restricted model	Binary logit	Truncated geometric	Geometric with zeros
INT	0.719 (0.231) [0.179]	1.089 (0.345)	0.627 (0.322)	0.429 (0.197)
AGE	− 0.0144 (0.0026) [0.0021]	− 0.0226 (0.0039)	− 0.0124 (0.0036)	− 0.0154 (0.0022)
EDUC	− 0.0232 (0.0127) [0.0104]	0.0032 (0.0194)	− 0.0536 (0.0175)	− 0.0205 (0.0107)
INCOME	−3.3E-6 (3.4E-6) [2.7E-6]	−7.6E-6 (5.0E-6)	−1.4E-6 (4.6E-6)	−3.7E-6 (2.8E-6)
MALE	0.372 (0.065) [0.051]	0.559 (0.100)	0.337 (0.089)	0.381 (0.054)
WHITE	0.154 (0.132) [0.096]	0.120 (0.191)	0.260 (0.191)	0.157 (0.122)
MARRIED	0.0106 (0.0754) [0.0574]	0.0125 (0.113)	− 0.00017 (0.104)	-0.0018 (0.0629)
ψ		—	—	− 0.354 (0.0401)
L, Λ	− 2736.81	− 1245.12	− 1436.60	− 2684.73

[a] Figures in parentheses are estimated asymptotic standard errors derived from the negative inverse Hessian of L evaluated at the ML estimates. Figures in square brackets are estimated asymptotic standard errors derived from the parameter covariance estimates obtained using the method proposed by White (1982) and Royal1 (1984).

alternative estimator where the implied variance/mean ratio is less than that of the geometric (e.g., Poisson) might be appropriate.

Table 6 presents the ML estimates of the intercept-only models. Among other things, table 6 demonstrates two points noted earlier: first, that the estimates of the intercept parameters in the truncated and the WZ models will be identical when only an intercept term is included; and, second, that the relationship $\psi = [1 - \exp(\alpha)]/[1 + \exp(\beta + \alpha)]$ obtains between the estimated parameters in the WZ and hurdle models.[30]

Tables 7 and 8 summarize the specification test results for the models with covariates included and the intercept-only models, respectively. In the first

[30] Considering the tea model as an example, and using the reparameterizations $\hat{\beta} = \hat{\beta}_T = 0.124$ and $\hat{\alpha} = \hat{\beta}_L - \hat{\beta}_T = -0.598$, where $\hat{\beta}_L$ and $\hat{\beta}_T$ are the logit and truncated-geometric intercepts, then $\hat{\psi} = 0.277 = (1 - \exp(-0.598))/(1 + \exp(-0.474))$.

Table 6

Estimation results: Intercept-only models.[a]

	Restricted model	Binary logit	Truncated geometric	Geometric with zeros
		COFFEE		
INT	0.995	1.032	0.982	0.982
	(0.027)	(0.052)	(0.031)	(0.031)
	[0.024]			
ψ				-0.0137
				(0.016)
L, Λ	−4105.35	−1094.00	−3010.99	−4105.00
		TEA		
INT	−0.201	−0.474	0.124	0.124
	(0.034)	(0.047)	(0.051)	(0.051)
	[0.038]			
ψ		—		0.211
				(0.027)
L, Λ	−2316.11	−1265.09	−1074.18	−2339.28
		MILK		
INT	0.104	0.390	−0.150	−0.150
	(0.032)	(0.047)	(0.044)	(0.044)
	[0.026]			
ψ		—		−0.289
				(0.039)
L, Λ	−2112.13	−1281.51	−1455.23	−2136.13

[a] Figures in parentheses are estimated asymptotic standard errors derived from the negative inverse Hessian of L evaluated at the ML estimates. Figures in square brackets are estimated asymptotic standard errors derived from the parameter covariance estimates obtained using the method proposed by White (1982) and Royal1 (1984).

rows are presented the results of White's IM test applied to the basic geometric models. As argued in section 3, the IM test can be viewed as an omnibus test for model misspecification. For both the models with covariates and the intercept-only models, the IM tests strongly suggest that the basic model is a misspecification of the dgp, as the null hypothesis of no misspecification is rejected in all but one instance at greater than the 0.9999 level.[31]

[31] When an intercept term is included in the X_t vectors, the presence of O-l dummy variables in X_t reduces the number of distinct upper triangular elements in $\nabla^2 L_t$ to at most $q = 0.5m(m+1) - d$, where L_t is the contribution of the tth observation to the restricted loglikelihood, m is the number of columns in X_t, and d is the number of O-l dummy variables. Since in the present application m = 7 and $d = 3$, the information matrix test statistics in the models with covariates are distributed χ^2_{25}. The method proposed by Lancaster (1984, eq. 6) is used to calculate the IM test statistics.

Table 7

Specification test results (models with covariates included).[a]

		COFFEE	*TEA*	*MILK*
Information matrix	[25]	266.130	56.1775	403.064
		(0.9999)	(0.9997)	(0.9999)
		Hurdle models		
Likelihood ratio	[7]	68.5451	77.0992	100.192
		(0.9999)	(0.9999)	(0.9999)
Score (Hessian)	[7]	64.1771	76.9506	108.685
		(0.9999)	(0.9999)	(0.9999)
Score (Gradient)	[7]	77.4802	89.7063	123.968
		(0.9999)	(0.9999)	(0.9999)
Score (White-Engle)	[7]	76.8902	87.0582	116.124
		(0.9999)	(0.9999)	(0.9999)
Hausman test: geometric		57.5672	70.5046	99.7153
vs. logit	[7]	(0.9999)	(0.9999)	(0.9999)
Hausman test: geometric		87.2395	75.8893	109.848
vs. truncated-geometric	[7]	((0.9999)	(0.9999)	(0.9999)
		With-zeros models		
Likelihood ratio	[1]	17.0983	66.1365	104.158
		(0.9999)	(0.9999)	(0.9999)
Score (Hessian)	Ill	17.4171	loo.113	83.8957
		(0.9999)	(0.9999)	(0.9999)
Score (Gradient)	[1]	18.7919	73.6177	115.890
		(0.9999)	(0.9999)	(0.9999)
Score (White-Engle)	[1]	18.3835	68.3445	100.599
		(0.9999)	(0.9999)	(0.9999)

“Figures in parentheses are $\Pr(\chi_q^2 < s)$, where s is the test statistic and q is the degrees of freedom of the test statistic; a value of 0.9999 signifies $\Pr(X_q^2 < s) \geq 0.9999$. Figures in square brackets are the degrees of freedom of the test statistics. The test statistics have asymptotic central χ_a^2 distributions under the null.

The results of the tests designed to test the restricted models against their corresponding hurdle variants are presented in rows 2-7 of tables 7 and 8. For each model, the range of the six test statistics is quite small, and in all cases except the intercept-only coffee model, rejection of the parameter restrictions specified under the null hypothesis is indicated.[32] Moreover, in each instance

[32]White (1982, p. 8) has noted that the use of the standard likelihood ratio test is not appropriate in cases where the probability densities that form the sample likelihood function are r&specified. In addition, the standard Hausman test approach uses estimates of the two covariance matrixes that are consistent under the null hypothesis of no misspecification; accordingly, no attempt was made to utilize alternative covariance estimators in calculating the Hausman tests. The covariance estimates used to construct the Hausman test statistics are the inverses of the matrixes in (A.1)–(A.3) in appendix A evaluated in the ML estimates. As shown in appendix A, these estimates guarantee that the difference of the covariance matrix estimates in (29) will be positive semidefinite as required for the Hausman test.

Table 8

Specification test results (intercept-only models).[a]

		COFFEE	TEA	MILK
Information matrix	[1]	40.8520 (0.9999)	25.8139 (0.9999)	208.268 (0.9999)
		Hurdle models		
Likelihood ratio	[1]	0.6916 (0.5964)	74.9872 (0.9999)	72.0001 (0.9999)
Score (Hessian)	[1]	0.6951 (0.5956)	14.8446 (0.9999)	71.7445 (0.9999)
Score (Gradient)	[1]	0.1616 (0.6172)	82.3494 (0.9999)	80.1355 (0.9999)
Score (White-Engle)	[1]	0.7512 (0.6139)	81.3942 (0.9999)	16.2101 (0.9999)
Hausman test: geometric vs. logit	[1]	0.6906 (0.5940)	70.5412 (0.9999)	69.1072 (0.9999)
Hausman test: geometric vs. truncated-geometric	[1]	0.1010 (0.6000)	74.5735 (0.9999)	71.1284 (0.9999)
		With-zeros models		
Likelihood ratio	[1]	0.6916 (0.5964)	74.9812 (0.9999)	72.0001 (0.9999)
Score (Hessian)	[1]	0.6976 (0.5964)	119.125 (0.9999)	58.8129 (0.9999)
Score (Gradient)	[1]	0.7616 (0.6172)	82.3494 (0.9999)	80.1355 (0.9999)
Score (White-Engle)	[1]	0.7484 (0.6130)	11.0294 (0.9999)	71.5301 (0.9999)

‘Figures in parentheses are $\Pr(\chi_q^2 < s)$, where s is the test statistic and q is the degrees of freedom of the test statistic; a value of 0.9999 signifies $\Pr(\chi_q^2 < s) \geq 0.9999$. Figures in square brackets are the degrees of freedom of the test statistics. The test statistics have asymptotic central χ_q^2 distributions under the null.

the score test statistics calculated using the White-Engle approach are smaller and larger than those calculated using the gradient outer product and negative Hessian, respectively, to estimate the information matrix.

It is interesting that the range of the test statistics for each model is relatively small: although each statistic has the same asymptotic distribution under the null hypothesis, the similarity of their finite-sample behavior when rejection of the null is favored was not anticipated ex ante.

The results of the tests of the basic models against the corresponding WZ specifications are presented in rows 8-11 of tables 7 and 8. Again the values of the test statistics fall within narrow ranges. Two results are particularly interesting here. First, in the covariates-included and intercept-only tea models, the score test calculated using the White-Engle method is smaller than those based on the gradient outer product and negative Hessian methods. Second, although still indicating rejection of the null hypothesis, the test statistics for the coffee model with covariates included are substantially smaller than those for the tea and milk models. In the intercept-only coffee model, none of the test statistics recommends rejection of the null at conventional confidence levels. Upon examination of the ratio of the estimates of ψ to their asymptotic standard errors in the coffee models, such results are not surprising.

5. Summary

This paper has explored the specification and testing of some variants on familiar count data models. The alternative specifications considered were termed hurdle and with-zero models, from which the familiar models were demonstrated to arise through parameter restrictions. Both alternatives were shown to allow for a degree of flexibility in model specification that is precluded by the basic model. In particular, it was seen that overdispersion and underdispersion could be accounted for by both alternatives. Score, Hausman, and information matrix tests for misspecification were proposed. The ideas were illustrated by estimating count data models of beverage consumption using survey microdata. In virtually all instances, the specification tests recommended rejection of the null hypothesis of no misspecification. For a given model, the different test statistics tended to behave quite similarly.

Appendix A

Since the ML logit and truncated-geometric estimators of the geometric model fail to utilize all sample information, their inefficiency relative to the ML geometric estimator follows immediately from the fact that the geometric estimator, which uses all sample information, is FIML. The following demonstration is illustrative. Let $L^{\mathrm{G}}, \Lambda^{\mathrm{L}}$, and Λ^{T} denote the loglikelihood functions of geometric, logit, and truncated-geometric models. Then

$$\Theta_{\mathrm{G}} = -L^{\mathrm{G}}_{\beta\beta} = \sum_{t\in\Omega} \left\{ (y_t+1)\exp(X_t\beta) \big/ [1+\exp(X_t\beta)]^2 \right\} X_t'X_t, \quad \text{(A.1)}$$

$$\Theta_{\mathrm{L}} = -\Lambda^{\mathrm{L}}_{\beta\beta} = \sum_{t\in\Omega} \left\{ \exp(X_t\beta) \big/ [1+\exp(X_t\beta)]^2 \right\} X_t'X_t, \quad \text{(A.2)}$$

$$\Theta_{\mathrm{T}} = -\Lambda^{\mathrm{T}}_{\beta\beta} = \sum_{t\in\Omega} \left\{ y_t\exp(X_t\beta) \big/ [1+\exp(X_t\beta)]^2 \right\} X_t'X_t. \quad \text{(A.3)}$$

Since $y_t \geq 0$ for all t, then $\Theta_{\mathrm{G}}, \Theta_{\mathrm{L}}$, and Θ_{T} are each positive semidefinite. It is easy to see from (A.1)–(A.3) that $(\Theta_G - \Theta_L) = \Theta_T$ and $(0, -\Theta_T) = \Theta_L$ if the Θ_i are all evaluated at the same β. The logit and truncated-geometric estimators are inefficient relative to the geometric estimator since both $(\Theta_G - 0,)$ and $(\Theta_G - \Theta_T)$, and therefore $(\Theta_L^{-1} - \Theta_G^{-1})$ and $(\Theta_T^{-1} - \Theta_G^{-1})$, are positive semidefinite.

The relative efficiencies of the logit and truncated-geometric estimators cannot in general be determined without knowledge of the sample (y,, X_t) values. In one extreme case where all y_t tend toward zeros and ones, it can be seen from (A.2) and (A.3) that $(0, -\Theta_T)$ becomes positive semidefinite, so that logit is efficient relative to truncated-geometric.[33] In another extreme instance where all y_t tend toward strictly positive integers, $(\Theta_T - 0,)$ becomes positive semidefinite, so that truncated-geometric is efficient relative to logit.

Appendix B

This appendix presents the gradient vectors and Hessian matrixes for the geometric hurdle and WZ models, which are used to construct the specification tests described in section 3 and implemented in section 4. For economy of space, the corresponding Poisson formulae are omitted here, but are available on request from the author.

The loglikelihood function of the geometric hurdle model (20) written in terms of the parameters $(\beta + \alpha) = \beta_1$ and $\beta = \beta_2$ is

$$\begin{aligned}\Lambda^{GH} = &\sum_{t \in \Omega_0} -\log\{1 + \exp[X_t(\beta + \alpha)]\} \\ &+ \sum_{t \in \Omega_1} X_t(\beta + \alpha) - \log\{1 + \exp[X_t(\beta + \alpha)]\} \\ &+ (y_t - 1)X_t\beta - y_t \log[1 + \exp(X_t\beta)]. \end{aligned} \tag{B.1}$$

The gradient and Hessian components are

$$\begin{aligned}\Lambda_\alpha^{GH} = &\sum_{t \in \Omega_0} (-\exp[X_t(\beta + \alpha)]/\{1 + \exp[X_t(\beta + \alpha)]\})X_t' \\ &+ \sum_{t \in \Omega_1} (1/\{1 + \exp[X_t(\beta + 41 \})X_t' \end{aligned} \tag{B.2}$$

[33] Even in this extreme case the geometric estimator remains efficient relative to the logit estimator, as the former uses information on the magnitude of the positive y, while the latter recognizes only their sign.

$$\underset{(\alpha=0)}{=} \sum_{t\in\Omega_0} \{-\exp(X_t\beta)/[1+\exp(X_t\beta)]\} X_t'$$

$$+ \sum_{t\in\Omega_1} \{1/[1+\exp(X_t\beta)]\} X_t', \tag{B.2'}$$

$$\Lambda_\beta^{\mathrm{GH}} = \sum_{t\in\Omega_0} \left(-\exp[X_t(\beta+\alpha)]/\{1+\exp[X_t(\beta+\alpha)]\}\right) X_t'$$

$$+ \sum_{t\in\Omega_1} \left[\left(1/\{1+\exp[X_t(\beta+\alpha)]\}\right)\right.$$

$$\left.+ \left([y_t - 1 - \exp(X_t\beta)]/[1+\exp(X_t\beta)]\right)\right] X_t' \tag{B.3}$$

$$\underset{(\alpha=0)}{=} \sum_{t\in\Omega} \{[y_t - \exp(X_t\beta)]/[1+\exp(X_t\beta)]\} X_t', \tag{B.3'}$$

$$\Lambda_{\alpha\alpha}^{\mathrm{GH}} = \Lambda_{\alpha\beta}^{\mathrm{GH}}$$

$$= \sum_{t\in\Omega} \left(-\exp[X_t(\beta+\alpha)]/\{1+\exp[X_t(\beta+\alpha)]\}^2\right) X_t'X_t \tag{B.4}$$

$$\underset{(\alpha=0)}{=} \sum_{t\in\Omega} \left\{-\exp(X_t\beta)/[1+\exp(X_t\beta)]^2\right\} X_t'X_t, \tag{B.4'}$$

$$\Lambda_{\beta\beta}^{\mathrm{GH}} = \sum_{t\in\Omega} \left[\left(-\exp[X_t(\beta+\alpha)]/\{1+\exp[X_t(\beta+\alpha)]\}^2\right)\right.$$

$$\left.- \left(y_t \exp(X_t\beta)/[1+\exp(X_t\beta)]^2\right)\right] X_t'X_t \tag{B.5}$$

$$\underset{(a=0)}{=} \sum_{t\in\Omega} \left\{-(y_t+1)\exp(X_t\beta)/[1+\exp(X_t\beta)]^2\right\} X_t'X_t, \tag{B.5'}$$

where the equalities in (B.2′)–(B.5′) hold under the restriction a = 0. The Hessian of (B.l) is

$$\Lambda_{\theta\theta}^{\mathrm{GH}} = \begin{bmatrix} \Lambda_{\alpha\alpha}^{\mathrm{GH}} & \Lambda_{\alpha\beta}^{\mathrm{GH}} \\ \Lambda_{\beta\alpha}^{\mathrm{GH}} & \Lambda_{\beta\beta}^{\mathrm{GH}} \end{bmatrix}. \tag{B.6}$$

Note that adding to and subtracting from the numerator of each term in the

Ω_1 summation in (B.2') the expression $[y_t - \exp(X_t\beta)]$, and using (B.3'), the non-zero elements of the score vector (B.2') can be expressed as

$$\Lambda_{\alpha}^{\mathrm{GH}}|_{\alpha=0} = \sum_{t\in\Omega_1} \left(\{y_t - [1+\exp(X_t\beta)]\}/\{1+\exp(X_t\beta)\}\right) X_t'. \tag{B.2''}$$

Eq. (B.2") corresponds to the non-zero elements of the score vector in the score test. Eq. (B.2") is also equivalent to the first-order conditions for ML estimation of a truncated-geometric model, or, alternatively, is the cross-product of explanatory variables and the (variance-normalized) residuals from the truncated-geometric model.[34]

Using the equality[35]

$$\begin{bmatrix} A & A \\ A & B \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1}B(B-A)^{-1} & -(B-A)^{-1} \\ -(B-A)^{-1} & (B-A)^{-1} \end{bmatrix},$$

the $(1,1)$ block of $(-\Lambda_{\theta\theta}^{\mathrm{GH}})^{-1}$ required to calculate the score test is given by

$$\left(-\Lambda_{\theta\theta}^{\mathrm{GH}}\right)^{11} = \left(-\Lambda_{\alpha\alpha}^{\mathrm{GH}}\right)^{-1}\left(-\Lambda_{\beta\beta}^{\mathrm{GH}}\right)\left(-\Lambda_{\beta\beta}^{\mathrm{GH}} + \Lambda_{\alpha\alpha}^{\mathrm{GH}}\right)^{-1}, \tag{B.7}$$

with all evaluations at $\alpha = 0$. Eqs. (B.4) and (B.5) show that $(-\Lambda_{\beta\beta}^{\mathrm{GH}} + \Lambda_{\alpha\alpha}^{\mathrm{GH}})$ is positive semidefinite.

The gradient and Hessian components for the geometric WZ loglikelihood function (24) are

$$\Lambda_{\psi}^{\mathrm{GZ}} = \sum_{t\in\Omega_0} \{\exp(X_t\beta)/[\psi\exp(X_t\beta)+1]\} - \sum_{t\in\Omega_1} (1-\psi)^{-1} \tag{B.8}$$

$$\underset{(\psi=0)}{=} -T_1 + \sum_{t\in\Omega_0} \exp(X_t\beta), \tag{B.8'}$$

$$\begin{aligned} \Lambda_{\beta}^{\mathrm{GZ}} = \sum_{t\in\Omega_0} & \left(\{\psi\exp(X_t\beta)/[\psi\exp(X_t\beta)+1]\}\right. \\ & \left. - \{\exp(X_t\beta)/[1+\exp(X_t\beta)]\}\right) X_t \\ & + \sum_{t\in\Omega_1} \{[y_t - \exp(X_t\beta)]/[1+\exp(X_t\beta)]\} X_t' \end{aligned} \tag{B.9}$$

[34] $E(Y_t) = [1 + \exp(X_t\beta)]$ in the truncated-geometric model, so that the residual is $y_t - [1 + \exp(X_t\beta)]$. The interpretation of (B.2") as a function of the residuals of the truncated variant of the parent model is the same as that given by Lin and Schmidt (1984) of the corresponding equation used in construction of their Tobit model score test.

[35] This assumes the arbitrary square matrixes A and B are symmetric and non-singular.

$$\underset{(\psi=0)}{=} \sum_{t \in \Omega} \{[y_t - \exp(X_t\beta)]/[1 + \exp(X_t\beta)]\} X_t', \tag{B.9'}$$

$$\Lambda_{\psi\psi}^{GZ} = \sum_{t \in \Omega_0} \left\{ -\exp(2X_t\beta)/[\psi\exp(X_t\beta)+1]^2 \right\}$$

$$+ \sum_{t \in \Omega_1} -(1-\psi)^{-2} \tag{B.10}$$

$$\underset{(\psi=0)}{=} -T_1 - \sum_{t \in \Omega_0} \exp(2X_t\beta), \tag{B.10'}$$

$$\Lambda_{\psi\beta}^{GZ} = \sum_{t \in \Omega_0} \left\{ \exp(X_t\beta)/[\psi\exp(X_t\beta)+1]^2 \right\} X_t \tag{B.11}$$

$$\underset{(\psi=0)}{=} \sum_{t \in \Omega_0} \exp(X_t\beta) X_t, \tag{B.11'}$$

$$\Lambda_{\beta\beta}^{GZ} = \sum_{t \in \Omega_0} \left\{ \psi\exp(X_t\beta)/[\psi\exp(X_t\beta)+1]^2 \right\} X_t'X_t$$

$$- \sum_{t \in \Omega} \left\{ (y_t+1)\exp(X_t\beta)/[1+\exp(X_t\beta)]^2 \right\} X_t'X_t \tag{B.12}$$

$$\underset{(\psi=0)}{=} \sum_{t \in \Omega} \left\{ -(y_t+1)\exp(X_t\beta)/[1+\exp(X_t\beta)]^2 \right\} X_t'X_t, \tag{B.12'}$$

where $T_1 = \#\Omega_1$, and the equalities in (B.8′)–(B.12′) hold under the restriction $\psi = 0$. Two points are noteworthy. First, unlike the geometric hurdle model, the non-zero elements of the score vector (B.8′) are not clearly interpretable as a function of residuals. Second, a simplification analogous to (B.7) is not apparent here, so that the full Hessian of $-\Lambda^{GZ}$ would have to be inverted to calculate the score test statistic.

References

Amemiya, T., 1984, Tobit models: A survey, Journal of Econometrics 24, 3-61.

Breusch, T. and A.R. Pagan, 1980, The Lagrange multiplier test and its application to model specification in econometrics, Review of Economic Studies 47, 239-253.

Cameron, A.C. and P.K. Trivedi, 1986, Econometric models based on count data: Comparisons and applications of some estimators and tests, Journal of Applied Econometrics 1, 29-53.

Cox, D.R., 1983, Some remarks on overdispersion, Biometrika 70, 269-274.

Cragg, J.G., 1971, Some statistical models for limited dependent variables with application to the demand for durable goods, Econometrica 39, 829-844.

Davidson, R. and J.G. MacKinnon, 1984a, Convenient specification tests for logit and probit models, Journal of Econometrics 25, 241-262.
Davidson, R. and J.G. MacKinnon, 1984b, Model specification tests based on artificial linear regressions, International Economic Review 25, 485-502.
Engle, R.F., 1984, Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in: Z. Grihches and M.D. Intriligator, eds. Handbook of econometrics, Vol. II (North-Holland, Amsterdam) 775-826.
Gourieroux, C., A. Montfort and A. Trognon, 1984a, Pseudo maximum likelihood methods: Theory, Econometrica 52, 681-700.
Gourieroux, C., A. Montfort and A. Trognon, 1984b, Pseudo maximum likelihood methods: Applications to Poisson models, Econometrica 52, 701-720.
Hausman, J.A., 1978, Specification tests in econometrics, Econometdca 46, 1251-1271.
Hausman, J.A., B. Hall and Z. Griliches, 1984, Econometric methods for count data with an application to the patents-R&D relationship, Econometrica 52, 909-938.
Hausman, J.A., B. Ostro and D. Wise, 1984. Air pollution and lost work, Working paper no. 1263 (National Bureau of Economic Research, Cambridge, MA).
Johnson, N.L. and S. Kotz, 1969, Distributions in statistics: Discrete distributions (Wiley, New York).
Lancaster, T., 1984, The covariance matrix of the information matrix test, Econometrica 52, 1051-1053.
Lee, L.-F., 1984a, Specification tests for Poisson regression models, Discussion paper no. 208 (Center for Economic Research, University of Minnesota, Minneapolis, MN).
Lee, L.-F., 1984b, Comment to tests of specification in econometrics, Econometric Reviews 3, 257-259.
Lin, T.-F. and P. Schmidt, 1984, A test of the Tobit specification against an alternative suggested by Cragg, Review of Economics and Statistics 66, 174-177.
Manning, W., L. Lillard and C.E. Phelps, 1983, Preventive medical care and its consequences (Rand Corporation, Santa Monica, CA).
McCullagh, P. and J.A. Nelder, 1983, Generalized linear models (Chapman and Hall, London).
Portney, P.R. and J. Mullahy, 1986, Urban air quality and acute respiratory illness, Journal of Urban Economics 20, 21-38.
Rosenzweig, M.R. and K.I. Wolpin, 1982, Governmental interventions and household behavior in a developing country: Anticipating the unanticipated consequences of social programs, Journal of Development Economics 10, 209-225.
Royall, R.M., 1984, Robust inference using maximum likelihood estimators, Working paper no. 549 (Department of Biostatistics, Johns Hopkins University, Baltimore, MD).
Ruud, P.A., 1984, Tests of specification in econometrics, Econometric Reviews 3, 211-242.
Stapleton, D.C. and D.J. Young, 1984, Censored normal regression with measurement error on the dependent variable, Econometrica 52, 737-760.
Terza, J.V., 1985, A Tobit-type estimator for the censored Poisson regression model, Economics Letters 18, 361-365.
U.S. Department of Commerce, 1982, Micro-data tape documentation for wave I and II of the national survey of personal health practices and consequences, NTIS publication no. PB83-104315.
White, H., 1982, Maximum likelihood estimation of r&specified models, Econometrica 50, l-25,
White, H., 1983, Corrigendum, Econometrica 51, 513.