



University of Essex

Department of Economics

## Discussion Paper Series

No. 661 October 2008

### Specification and Testing of Models Estimated by Quadrature

Geert Dhaene and J. M. C. Santos Silva

Note : The Discussion Papers in this series are prepared by members of the Department of Economics, University of Essex, for private circulation to interested readers. They often represent preliminary reports on work in progress and should therefore be neither quoted nor referred to in published work without the written consent of the author.

# Specification and testing of models estimated by quadrature\*

Geert Dhaene<sup>†</sup>      J.M.C. Santos Silva<sup>‡</sup>

April 10, 2010

## Abstract

This paper proposes a test to check the specification of models with unobserved individual effects integrated out by quadrature and also a simple way of increasing the flexibility of this type of model. The results of a Monte Carlo study and an application using a well-known data set illustrate the finite sample properties of the proposed methods and their implementation in practice.

*JEL classification code:* C12, C15, C23.

*Key words:* Bootstrap, Random effects, Transformations to normality.

## 1. INTRODUCTION AND SET-UP

Models where random effects are integrated out using quadrature are becoming ever more popular in applied statistics and econometrics. Because these models are based on a parametric specification of the distribution of the unobservable individual effects, there may be concern about potential departures from the maintained distributional

---

\*We are grateful to two anonymous referees and to Marcus Chambers, Daniel Dias, Gordon Kemp and Paulo Parente for helpful comments and discussions. Santos Silva also gratefully acknowledges financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

<sup>†</sup>K.U. Leuven, Department of Economics. E-mail: geert.dhaene@econ.kuleuven.be.

<sup>‡</sup>Department of Economics, University of Essex and CEMAPRE. E-mail: jmcass@essex.ac.uk.

assumptions. This paper proposes a simple procedure to test the specification of the distribution of the random effects and also an easy way to make this kind of model more flexible.

We consider fully parametric models for which the  $i$ -th observation has a contribution to the likelihood function specified as

$$L_i = \int f(y_i|x_i'\beta + \sigma\varepsilon_i) g(\varepsilon_i) d\varepsilon_i, \quad (1)$$

where  $y_i$  is the variate of interest;  $x_i$  is a vector of covariates with its first element equal to 1;  $\beta$  is a conformable vector of unknown parameters;  $\varepsilon_i$  is an unobservable individual effect; and  $\sigma > 0$  is an unknown parameter. As usual, it is supposed that the researcher is willing to specify  $f(y_i|\cdot)$ , the conditional density of  $y_i$ , and  $g(\varepsilon_i)$ , the marginal density of  $\varepsilon_i$ , which is assumed to be independent of the covariates and such that  $G(e) = \int_{-\infty}^e g(z) dz$  is strictly increasing in  $e$ .

A standard example of models of this form is the Poisson log-normal model (see Hinde, 1982, and Winkelmann, 2008), but the procedures developed here can easily be applied to more complex models, such as the random effects panel probit model (Wooldridge, 2002, p. 485) and the mixed multinomial logit model (McFadden and Train, 2000).

The key observation underlying our results is that replacing  $g(\cdot)$  with some other density is equivalent to transforming  $\varepsilon_i$  monotonically and keeping  $g(\cdot)$ . To see this, let  $h(\cdot)$  be the true density of  $\varepsilon_i$  and denote the corresponding distribution function by  $H(e) = \int_{-\infty}^e h(z) dz$ , which is assumed to be strictly increasing. The correctly specified model then has likelihood contributions of the form

$$L_i^* = \int f(y_i|x_i'\beta + \sigma\varepsilon_i) h(\varepsilon_i) d\varepsilon_i = \int f(y_i|x_i'\beta + \sigma\varepsilon_i) dH(\varepsilon_i).$$

Now, because  $u_i = H(\varepsilon_i)$  is uniformly distributed on  $[0, 1]$ , the variable  $G^{-1}(u_i) = G^{-1}(H(\varepsilon_i)) = m^{-1}(\varepsilon_i) = \zeta_i$  has density  $g(\cdot)$  and, by changing the variable of integration  $H(\varepsilon_i) = G(\zeta_i)$ , the likelihood contributions can be written as

$$L_i^* = \int f(y_i|x_i'\beta + \sigma m(\zeta_i)) g(\zeta_i) d\zeta_i = \int f(y_i|x_i'\beta + \sigma m(\varepsilon_i)) g(\varepsilon_i) d\varepsilon_i, \quad (2)$$

which provides a generalisation of (1) and an alternative against which this model can be tested. Indeed, the correct specification of the distribution of  $\varepsilon_i$  can be checked by testing  $H_0 : m(\varepsilon_i) = \varepsilon_i$ . If the null is rejected, the original model can be made more flexible by replacing  $\varepsilon_i$  with some approximation to  $m(\varepsilon_i)$ .

Except for very special cases in which the integral in (1) has a closed form,  $L_i$  has to be evaluated using some form of numerical integration. Whatever the type of numerical integration used, such as Monte Carlo methods or Gaussian quadrature, the model that is actually estimated has individual contributions to the likelihood function of the form

$$Q_i = \sum_{j=1}^q f(y_i | x_i' \beta + \sigma e_j) w_j, \quad (3)$$

where  $e_1, \dots, e_q$  is a set of appropriately chosen abscissas and  $w_1, \dots, w_q$  are the corresponding weights (see, e.g. Judd, 1998, for details). Naturally, the abscissas and the weights are chosen as a function of  $g(\cdot)$  so as to make  $Q_i$  a close approximation of  $L_i$ . The finite-mixture analogue of  $L_i^*$  is given by

$$Q_i^* = \sum_{j=1}^q f(y_i | x_i' \beta + \sigma m(e_j)) w_j.$$

In the next section, we give details on how  $H_0$  can be tested and develop flexible alternatives to (1) based on (2).

## 2. MAIN RESULTS

### 2.1. The proposed test

We propose to check the correct specification of (1) by testing  $Q_i$  against  $Q_i^*$ . This test will not only check the assumptions about the distribution of the unobservables but also provide a test for other assumptions underlying the specification of (3). Two particular cases are worth considering. First, there is a degree of duality between the misspecification of  $g(\cdot)$  and the misspecification of  $f(y_i | \cdot)$ . Indeed, by comparing (1) and (2), we can see that the misspecification of  $g(\cdot)$  can be interpreted as a misspec-

ification of the way the unobserved individual effect enters  $f(y_i|\cdot)$ .<sup>1</sup> Therefore, any appropriate test is likely to have power not only against misspecification of the mixing distribution but also against misspecification of the assumed conditional distribution of  $y_i$ , given  $x_i$  and  $\varepsilon_i$ . Second,  $Q_i$  can be severely misspecified even if  $L_i$  is not. Indeed, if the quadrature method used is too crude,  $Q_i$  will be a poor approximation of  $L_i$ . Therefore, a test for the correct specification of (3), which is the model actually estimated, will also provide a check on the quality of the numerical approximation used in the estimation.

Several avenues can be pursued in order to obtain an appropriate test. If the researcher is willing to specify a function  $m(\cdot)$  that depends on a small number of parameters and that has the identity function as a special case, standard tests can be used to check the validity of the restrictions. For example, the researcher may follow MacKinnon and Magee (1990) and consider a test against alternatives of the form  $m(z) = l(\theta z)/\theta$ , where  $l(\cdot)$  is such that  $l(0) = 0$ ,  $l'(0) = 1$  and  $l''(0) \neq 0$ , with  $l'$  and  $l''$  denoting the first two derivatives of  $l(\cdot)$ .<sup>2</sup>

However, if one prefers to be agnostic about possible departures from the specified distribution of the individual effects, a test can be based on a Taylor series expansion of  $m(\varepsilon_i)$ , as in the RESET test (see Ramsey, 1969, and also Cramer and Ridder, 1988). In practice, the order of the polynomial used to construct the test has to be chosen by the researcher. The simulation evidence provided in Section 3 suggests that a one-degree-of-freedom test based on a quadratic approximation to  $m(\varepsilon_i)$ , which is locally equivalent to testing  $\theta = 0$  when the alternative is  $m(z) = l(\theta z)/\theta$ , has good power in a variety of circumstances. Nevertheless, for greater generality, we consider

---

<sup>1</sup>The reverse, however, is not always true. For example, if  $f(y_i|\cdot)$  depends on  $x_i$  in a complex way, it may not be possible to write a correctly specified model based on a single-index specification of  $f(y_i|\cdot)$  and a change of the specification of  $g(\cdot)$ .

<sup>2</sup>Note that  $m(z) = l(\theta z)/\theta$  has the identity function as a limiting case as  $\theta$  passes to zero.

a cubic approximation of the form

$$m(\varepsilon_i) = m(\kappa) + (\varepsilon_i - \kappa) m'(\kappa) + (\varepsilon_i - \kappa)^2 \frac{m''(\kappa)}{2} + (\varepsilon_i - \kappa)^3 \frac{m'''(\kappa)}{6} + R,$$

where  $\kappa$  is the point around which the expansion is made,  $R$  is the remainder term, and  $m'$ ,  $m''$  and  $m'''$  denote the first three derivatives of  $m(\cdot)$ .

Using this expansion,  $L_i^*$  can be approximated by

$$L_i^{**} = \int f(y_i | x_i' \beta + \delta_1 \varepsilon_i + \delta_2 \varepsilon_i^2 + \delta_3 \varepsilon_i^3) g(\varepsilon_i) d\varepsilon_i,$$

where the constants in the Taylor series expansion are subsumed in the intercept of the model and  $\delta_1 \propto \sigma(m'(\kappa) - \kappa m''(\kappa) + 0.5 \kappa m'''(\kappa))$ ,  $\delta_2 \propto \sigma(m''(\kappa) - \kappa m'''(\kappa))$ , and  $\delta_3 \propto \sigma m'''(\kappa)$ . When  $\delta_2 = \delta_3 = 0$ , then  $L_i^{**}$  reduces to  $L_i$  with  $\delta_1 = \sigma$ . Therefore,  $\delta_1$  provides no information on the adequacy of the specification of  $L_i$ . By contrast, whenever  $\delta_2$  or  $\delta_3$  are non-zero,  $m(\cdot)$  is not the identity function. Therefore, rejection of  $H_0' : \delta_2 = \delta_3 = 0$  implies the rejection of  $H_0 : m(\varepsilon_i) = \varepsilon_i$ .

To perform the test when the model is estimated by quadrature, the significance of  $\delta_2$  and  $\delta_3$  has to be checked in the finite-mixture analogue of  $L_i^{**}$ , which is

$$Q_i^{**} = \sum_{j=1}^q f(y_i | x_i' \beta + \delta_1 e_j + \delta_2 e_j^2 + \delta_3 e_j^3) w_j. \quad (4)$$

Because the model defined by  $Q_i^{**}$  is estimated by maximum likelihood, the standard trilogy of testing principles is available to check the significance of  $\delta_2$  and  $\delta_3$ . In what follows, we focus on the score test because it has the advantage of not requiring the estimation of the model under the alternative, which can be cumbersome. Moreover, if the researcher prefers to consider the family of alternatives  $m(z) = l(\theta z) / \theta$ , the use of the score test for  $\theta = 0$  avoids the need to specify  $l(\cdot)$ .

For this kind of model, the score vector and any estimate of its variance matrix are highly nonlinear functions of the data, so one should not expect the asymptotic  $\chi^2$  distribution to be a good approximation of the finite sample distribution of the score statistic, even in relatively large samples. Therefore, it is advisable to implement the

test with critical values estimated by the parametric bootstrap (see, e.g. Beran, 1988, and Horowitz, 1994 and 2001), which are second-order accurate.

One final issue must be mentioned. Even when  $L_i$  is correctly specified, there is a non-zero probability in finite samples that the estimate of  $\sigma$  is equal to zero. In these cases, it is obviously not possible to test the correct specification of the distribution of the unobserved individual effects. Therefore, the test we propose is conditional on the estimate of  $\sigma$  being positive. Of course, this will not affect the asymptotic distribution of the proposed test statistic but will have an impact on its finite sample behaviour.

## 2.2. More flexible models

In many situations – for example, if the null is rejected – the researcher may want to estimate a more flexible model, like the one defined in (2). Given the lack of information on the appropriate transformation to use, one possibility is to approximate  $m(\cdot)$  by a cubic polynomial, as in  $L_i^{**}$ . However, if the estimated polynomial is not one-to-one, the model cannot be interpreted as a generalisation of (1) obtained by considering a more flexible distribution of the individual effects. Therefore, although using a polynomial in  $\varepsilon_i$  is possibly the easiest way to generalise (1), it is of interest to consider other choices of  $m(\cdot)$ .

Even if  $g(\cdot)$  is not the standard normal density, guidance about the choice of  $m(\cdot)$  can be obtained from the vast literature on transformations to normality. In particular,  $m(\cdot)$  can be defined as any of the transformations proposed to achieve normality, such as those introduced by Box and Cox (1964) and MacKinnon and Magee (1990), or their inverses.

As an illustration of this approach, consider the case in which  $g(\cdot)$  is the standard normal density and the researcher wants to allow for excess kurtosis while maintaining symmetry. In this case, the  $\sinh^{-1}$  transformation studied by MacKinnon and Magee (1990) and Burbidge, Magee and Robb (1988) may be used to obtain

$m(z) = \sinh(\theta z)/\theta$ , which has the identity function as the limiting case when  $\theta$  passes to zero. If the symmetry assumption is considered too restrictive, this transformation can be generalised by including a location parameter as in MacKinnon and Magee (1990, p. 325).

Finally, if the researcher wants to relax the assumption that the marginal density of  $\varepsilon_i$  is independent of the covariates, the parameters of  $m(\cdot)$  can be written as functions of the covariates.<sup>3</sup> However, unless a parsimonious parameterization is adopted, estimation of the effects of the covariates on the parameters of  $m(\cdot)$  is likely to be too noisy, except in very rich data sets.

### 3. FINITE SAMPLE PROPERTIES AND AN APPLICATION

In a classic paper, Cameron et al. (1988) studied the relation between the demand for health care and health insurance. They used a sample of 5190 single individuals from the 1977-78 Australian Health Survey to estimate count data models (NegBin1) for various measures of the demand for health care. We use the same data here to study the finite sample properties of the proposed test and to apply the test and the suggested flexible models in an empirical setting. We focus on models for the variable NONPRESC, which measures the number of non-prescribed medications used in the two days preceding the survey. The regressors used are as in Cameron et al. (1988), and are described in Table 1.

---

<sup>3</sup>It is worth noting that, by allowing  $m(\cdot)$  to be a function of  $x_i$ , it is also possible to test the assumption that the density of  $\varepsilon_i$  is independent of the covariates.



Table 1: Description of the regressors

---

SEX	1 if female
AGE	Age in years divided by 100
INCOME	Annual income in Australian dollars divided by 1000
LEVYPLUS	1 if covered by private health insurance for private patient in public hospital
FREEPOOR	1 if covered by government due to low income, recent immigrant, unemployed
FREEOTHER	1 if covered by government due to old age, disability pension, invalid veteran or family of deceased veteran
ILLNESS	Number of illnesses in past 2 weeks, top coded at 5
ACTIVDAYS	Number of days of reduced activity in past 2 weeks due to illness or injury
GHQ	General health questionnaire score using Goldberg’s method
LIMCHRON	1 if chronic condition(s) and limited in activity
NLIMCHRON	1 if chronic condition(s) but not limited in activity

---

### 3.1. Simulations

We ran three sets of Monte Carlo experiments, all centred around the Poisson log-normal model. The first was run with data generated under the null to study level distortions and the other two with data generated under deviations from the null to study the power of the test. Each experiment was run with 10000 replications.

#### 3.1.1. Level distortions

For the experiments under the null, we generated data from a  $q$ -point finite mixture model, which can be seen as a Gauss-Hermite approximation of a Poisson log-normal model. In particular,  $y_i$  was generated as

$$y_i | x_i, \varepsilon_i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = \exp(x_i' \beta + \sigma \varepsilon_i), \quad (i = 1, \dots, n), \quad (5)$$

where  $x_i' \beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$  and  $(x_{1i}, x_{2i}) \sim \text{EDF}(\text{INCOME}, \text{LEVYPLUS})$ .<sup>4</sup> The individual effect  $\varepsilon_i$  was drawn, independently of  $x_i$ , from the  $q$ -point distribution defined by

$$\Pr[\varepsilon_i = e_j] = \frac{w_j}{\sqrt{\pi}}, \quad e_j = \sqrt{2} z_j, \quad (j = 1, \dots, q), \quad (6)$$

---

<sup>4</sup>The variables INCOME and LEVYPLUS were chosen because they are fairly representative of the kind of regressors used in applications.

where  $z_1, \dots, z_q$  and  $w_1, \dots, w_q$  are the abscissas and weights of the  $q$ -point Gauss-Hermite integration rule (see, e.g., Abramowitz and Stegun, 1972, p. 924). All draws of  $(y_i, x_i, \varepsilon_i)$  were independent across observations, replications and design points. We set  $\beta = (-.5, .5, -.25)'$ ,  $\sigma = .5$ ,  $q = 10$  and  $n = 250, 1000, 4000$ . These choices ensure that  $y_i$  has a realistic range of variation.

The parameters  $\beta$  and  $\sigma$  were estimated as  $\hat{\beta}$  and  $\hat{\sigma}$  by maximizing the log-likelihood

$$\sum_{i=1}^n \log \left( \frac{1}{y_i!} \sum_{j=1}^q \exp(-\exp(x_i' \beta + \sigma e_j) + y_i (x_i' \beta + \sigma e_j)) \frac{w_j}{\sqrt{\pi}} \right), \quad (7)$$

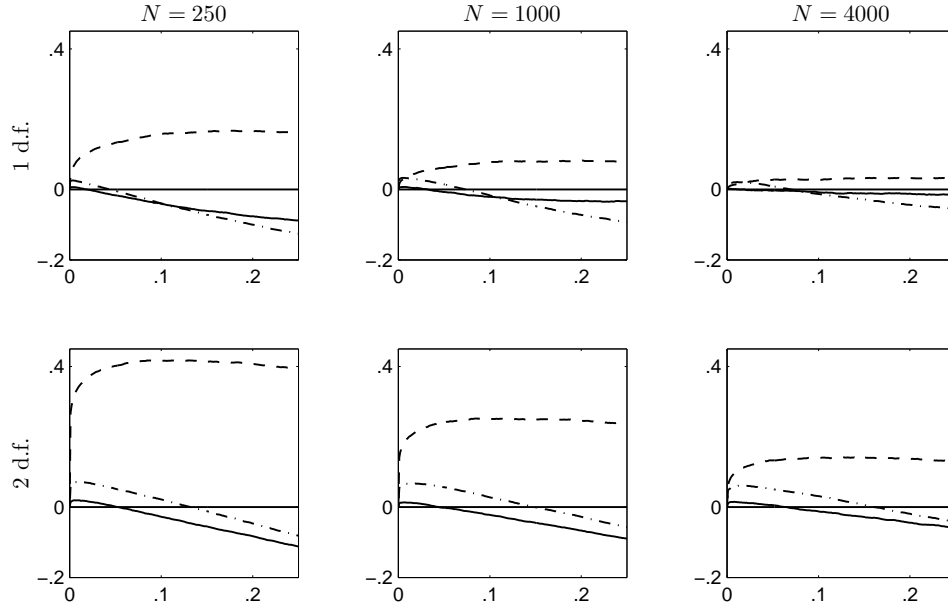
corresponding to (5)–(6). One may view (7) as a  $q$ -point Gauss-Hermite approximation of the log-likelihood of the Poisson log-normal model in which  $\varepsilon_i \sim N(0, 1)$  instead of (6) (see, e.g. Winkelmann, 2008). Conditional on  $\hat{\sigma} > 0$ , we computed one-d.f. and two-d.f. score test statistics for testing  $\delta_2 = 0$  and  $\delta_2 = \delta_3 = 0$  in the extended model (4).<sup>5</sup> We considered three versions of the statistics based on the Hessian, the OPG, and the robust “sandwich” estimator of the covariance matrix. Denote these as  $S_H$ ,  $S_O$  and  $S_R$ , respectively. For each statistic, say  $S$ , we computed the asymptotic  $p$ -value from the  $\chi^2$  distribution and the  $p$ -value from the parametric bootstrap distribution of  $S$ . The latter was estimated as the EDF of  $S(Y_b, X)$ ,  $b = 1, \dots, B$ , where  $S(Y_b, X)$  is  $S$  computed from  $(Y_b, X)$  conditional on  $\hat{\sigma}_b > 0$ ;  $X$  is the covariate data set in that replication;  $Y_b$  is the  $b^{\text{th}}$  bootstrap data set on the dependent variable (generated as above but with  $X$  held fixed and  $\hat{\beta}, \hat{\sigma}$  replacing  $\beta, \sigma$ ); and  $B$  is the number of bootstrap draws. We set  $B = 999$ .

The results are presented in the form of  $p$ -value discrepancy plots (Davidson and MacKinnon, 1998), which graph the level error of a test against its nominal level. Deviations from the zero line indicate level distortions: a test overrejects (or underrejects) at a given level when the ordinate of the  $p$ -value discrepancy plot is positive (or negative). Figure 1 shows  $p$ -value discrepancy plots for the tests with asymptotic  $p$ -values. All three versions of the test are severely level-distorted with the two-d.f. test

---

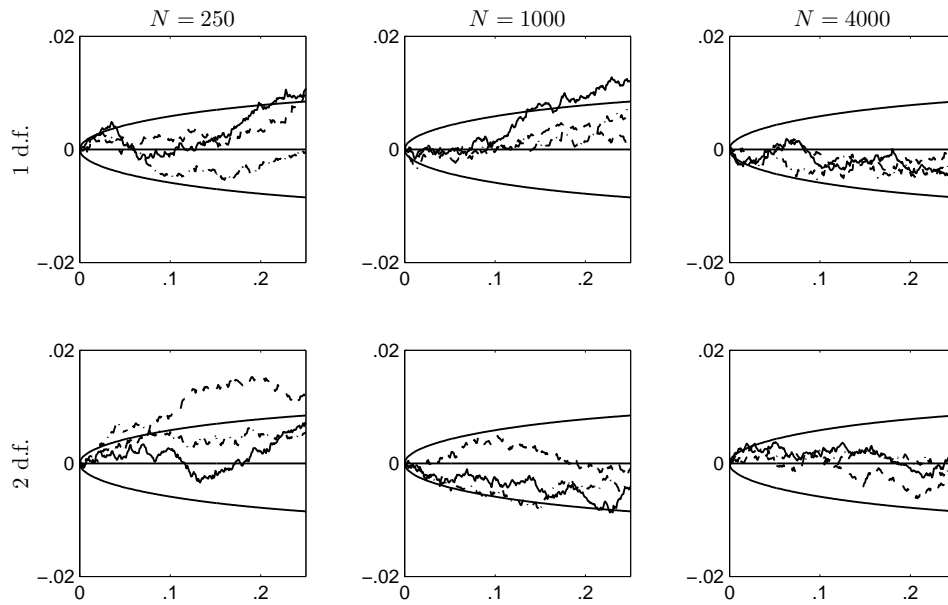
<sup>5</sup>In the one-d.f. test,  $\delta_3$  is set to 0 and this restriction is not tested.

Figure 1: Asymptotic  $p$ -value discrepancy plots



x-axis: nominal level; y-axis: level error; solid: zero line and  $S_H$ ; dashed:  $S_O$ ; dashed-dotted:  $S_R$ .

Figure 2: Bootstrap  $p$ -value discrepancy plots



x-axis: nominal level; y-axis: level error; solid: zero line,  $\pm 1.96\sqrt{x(1-x)}/10^4$  and  $S_H$ ; dashed:  $S_O$ ; dashed-dotted:  $S_R$ .

being the most level-distorted. Therefore, the tests should generally not be used with  $\chi^2$  critical values. Figure 2 shows  $p$ -value discrepancy plots (the jagged curves) for the tests with bootstrap  $p$ -values. All three tests are now nearly level-correct. Note that the jagged appearance of the curves is a consequence of the level distortions being small. We added pointwise 95% error bounds  $\pm 1.96\sqrt{x(1-x)}/10^4$  for an unbiased test (the smooth curves) as an indication of experimental randomness.

We also experimented with  $\sigma = .25, .75$  and  $q = 4, 20$ .<sup>6</sup> The results changed very little with  $q$  but they did change with  $\sigma$ . The level distortions (with and without bootstrap) decreased as  $\sigma$  increased, presumably due to the conditioning on  $\hat{\sigma} \neq 0$ . Nevertheless, even when  $\sigma = .75$ , the level distortions of the two-d.f. asymptotic tests remained large. In contrast, even when  $\sigma = .25$ , the level distortions were mild when the bootstrap was used; in particular, the tests based on  $S_O$  were nearly level-correct, while those based on  $S_H$  and  $S_R$  erred on the conservative side but had a rejection rate of at least 2.5% uniformly over the design when the nominal level is 5%.

In view of these results, we conclude that the use of bootstrap critical values makes the test nearly level-correct or leaves, at worst, only mild level distortions. We also note that, interestingly, the results parallel those obtained in the very different context of the information matrix test (see, e.g., Horowitz, 1994 and 2001; Davidson and MacKinnon, 1998; Horowitz and Savin, 2000).

### 3.1.2. Power

We considered two types of deviations from the null. In the first, the mixing distribution (6) was altered; in the second, excess zeroes were introduced.

In the first set of alternatives, data were simulated as under the null above but now with  $\varepsilon_i$  generated as

$$\varepsilon_i = \frac{\eta_i - \nu + (1 + \eta_i)u_i}{\sqrt{1 + 2\nu + 3\nu^2}}, \quad \eta_i \sim \text{Exp}(\nu), \quad u_i \sim N(0, 1), \quad (8)$$

---

<sup>6</sup>We ran a full factorial design over  $\sigma = .25, .5, .75$ ;  $q = 4, 10, 20$ ;  $n = 250, 1000, 4000$ , and  $B = 99$ . The results are available on request.

where  $\eta_i$  and  $u_i$  are independent and  $\text{Exp}(\nu)$  is the exponential distribution with mean  $\nu$ . For any  $\nu \geq 0$ ,  $\varepsilon_i$  has zero mean and unit variance, and its skewness and kurtosis increase in  $\nu$ . We set  $\sigma = .5$  and  $\nu = .5$ , implying that  $\varepsilon_i$  has skewness .71 and kurtosis 5.73. We used  $q = 10$  quadrature points to compute estimates (maximizing (7), as before) and score statistics, and  $B = 999$  to compute bootstrap  $p$ -values.

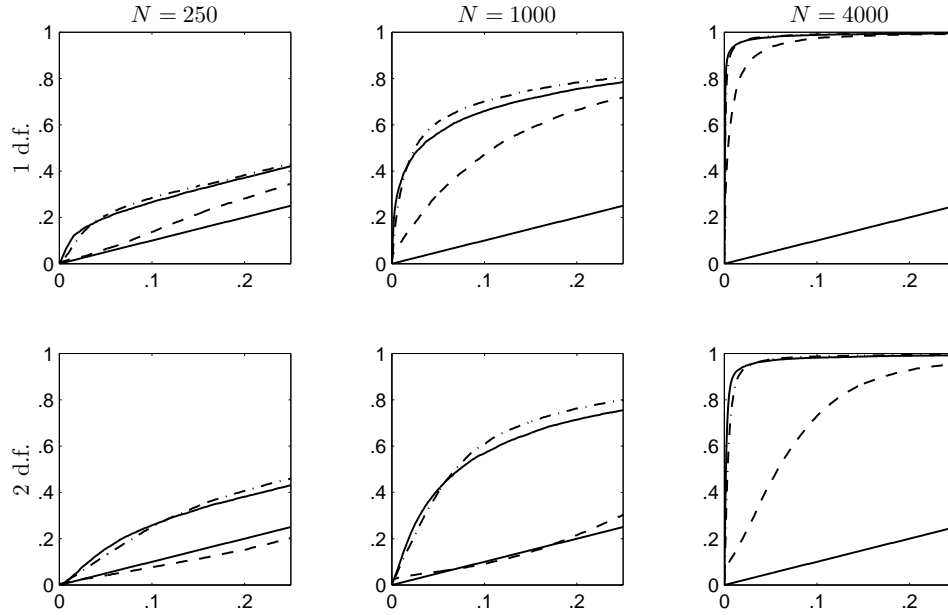
Figure 3 displays power as a function of nominal level. The tests based on  $S_O$  always have the least power (sometimes below the level), regardless of the level. The tests based on  $S_H$  and  $S_R$  have approximately the same power, with the one-d.f. tests being more powerful than the two-d.f. tests. The power increases rapidly with  $n$ . The results for  $\sigma = .25$ ;  $q = 4, 20$ ;  $\nu = 0, 1$  (available on request) exhibit the same patterns. Again, altering  $q$  gives nearly identical results. When  $\sigma = .25$ , the power drops uniformly, as expected, but remains important when  $\nu$  and  $n$  are sufficiently large. Also as expected, the power increases in  $\nu$ . When  $\nu = 0$ , the tests have virtually no power in excess of the level, which indicates that the Gauss-Hermite approximation  $Q_i$  is very close to  $L_i$ .<sup>7</sup>

To gain some insight into the ability of the test to detect other sorts of misspecification, we also ran simulations with zero-inflated data. Here, data were first simulated as above, but with  $\nu = 0$ . On obtaining  $y_i$ , misspecification was introduced by setting  $y_i$  to zero with probability  $\pi$  regardless of the value of  $y_i$  that was obtained. This is equivalent to having  $\varepsilon_i \sim N(0, 1)$  with probability  $1 - \pi$ , and  $\varepsilon_i = -\infty$  with probability  $\pi$ . Figure 4 presents power plots for  $\pi = .2$  with, as before,  $\sigma = .5$ ,  $q = 10$ , and  $B = 999$ . Again, the one-d.f. tests are more powerful than the two-d.f. tests, but now the tests based on  $S_R$  have the least power. At conventional levels, the one-d.f. test based on  $S_O$  is somewhat more powerful than the one based on  $S_H$  for  $n$  up to 1000.

---

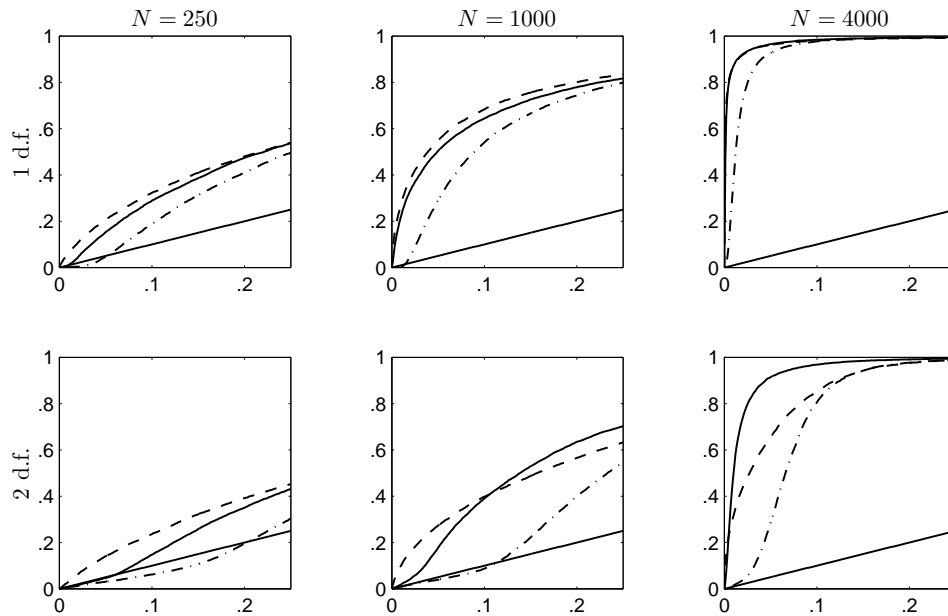
<sup>7</sup>This changes, however, when  $\sigma$  is very large and  $q$  very small. For example, when  $\nu = 0$ ,  $q = 4$ ,  $\sigma \geq .75$  and  $n \geq 1000$ , the tests have non-trivial power.

Figure 3: Power plots, non-normal mixture



Non-normal mixture data with  $\sigma = \nu = .5$ ; x-axis: nominal level; y-axis: power; solid:  $45^\circ$  line and  $S_H$ ; dashed:  $S_O$ ; dashed-dotted:  $S_R$ .

Figure 4: Power plots, zero-inflated data



Normal-mixture data with  $\sigma = .5$ , then zero-inflated; x-axis: nominal level; y axis: power; solid:  $45^\circ$  line and  $S_H$ ; dashed:  $S_O$ ; dashed-dotted:  $S_R$ .

Considering the sample sizes typically available when this sort of model is estimated, the results of this and the preceding experiments suggest that the proposed specification test, especially the one-d.f. version based on  $S_H$ , is an interesting additional tool that can be used in the evaluation of models estimated by quadrature methods.

### 3.2. Empirical illustration

Table 2 contains the estimation results for the NegBin1 regression model as specified by Cameron et al. (1988) and for the Poisson log-normal model estimated using Gauss-Hermite integration with  $q = 10$ . According to the value of the maximized log-likelihood, the latter model fits the data somewhat better than does the former.

Despite the improvement in the fit provided by the Poisson log-normal model, it is important to test if this specification is adequate for this data set. The two-d.f. score statistics in the Poisson log-normal model have values  $S_H = 12.44$ ,  $S_O = 6.85$  and  $S_R = 26.91$  with associated  $p$ -values based on  $B = 9999$  bootstrap draws equal to .027, .226 and .049. The corresponding one-d.f. test statistics are  $S_H = 8.74$ ,  $S_O = 6.02$ , and  $S_R = 16.11$  with associated bootstrap  $p$ -values equal to .004, .044, and .005, respectively. Recalling that the one-d.f. tests typically have better power than the two-d.f. versions, the results provide clear evidence against the hypothesis that this specification is appropriate.

In view of this finding, it is of interest to make the model more flexible by considering one of the specifications suggested in Section 2. The last two columns in Table 2 display the results obtained by estimating the model – labelled flexible Poisson log-normal – defined in (2), with  $m(z) = \sinh(\theta z)/\theta$  (again using Gauss-Hermite integration with  $q = 10$ ).

Since all the models estimated here have the same specification of the mean function, it is not surprising to find that the estimates of the slope parameters do not

vary much across the models.<sup>8</sup> However, judging by the value of the maximized log-likelihood, the flexible Poisson log-normal fits the data better than do the two competitors considered here.

Table 2: Estimates

	NegBin1		Poisson log-normal		Flexible Poisson log-normal	
	estimates	s.e.	estimates	s.e.	estimates	s.e.
INTERCEPT	-1.021	.207	-2.646	.206	-2.612	.203
SEX	.239	.057	.251	.060	.254	.059
AGE	4.816	1.055	4.904	1.123	5.084	1.103
AGE <sup>2</sup>	-6.124	1.181	-6.181	1.244	-6.434	1.226
INCOME	.055	.080	.092	.083	.075	.083
LEVYPLUS	-.045	.065	-.043	.070	-.046	.068
FREEPOOR	-.077	.142	-.037	.138	-.106	.152
FREEOTHER	-.285	.101	-.288	.105	-.303	.104
ILLNESS	.205	.020	.212	.022	.210	.021
ACTIVDAYS	-.005	.008	.000	.009	.001	.008
GHQ	.028	.012	.030	.013	.032	.012
LIMCHRON	.014	.091	.000	.098	.013	.096
NLIMCHRON	.135	.061	.153	.063	.167	.063
$\delta$	3.435	.331	—	—	—	—
$\sigma$	—	—	.790	.035	.618	.094
$\theta$	—	—	—	—	.525	.172
log likelihood	-3929.09		-3923.15		-3919.31	

*Note:* OPG-based standard errors.

In order to check if the differences in the fit of the models are significant,  $\chi^2$  goodness-of-fit tests (see Cameron and Trivedi, 1998, pp. 155-157) were performed. Essentially, these tests check whether the predicted probabilities of certain counts are compatible with what is observed in the data. The results of the  $\chi^2$  goodness-of-fit tests are summarized in Table 3. Using the 5% level, the results in this table show that only the flexible Poisson log-normal model passes the three goodness-of-fit tests.

<sup>8</sup>This does not imply that the estimated partial effects are similar: the partial effect of AGE differs by up to 26% across the different models, and the differences for the partial effect of NLIMCHRON can be larger than 40%.



Table 3: Goodness-of-fit tests

	NegBin1		Poisson log-normal		Flexible Poisson log-normal		
	d.f.	statistic	$p$ -value	statistic	$p$ -value	statistic	$p$ -value
Count of 0	1	9.47	.002	6.05	.014	1.46	.226
Counts of 0 to 2	3	14.12	.003	11.16	.011	2.35	.503
Counts of 0 to 5	6	15.46	.017	11.54	.073	8.50	.204

In conclusion, although the Poisson log-normal model represents a substantial improvement over the NegBin1 model, both the specification test proposed in this paper and the goodness-of-fit tests suggest that it is misspecified for this data set. The flexible Poisson log-normal model seems to provide a more accurate description of the data.

#### 4. CONCLUDING REMARKS

Models estimated using quadrature methods rely on strong distributional assumptions that are often untested. One of the reasons why departures from these assumptions are not tested more often may be because no simple test for these hypotheses has been suggested in the literature. The test suggested in this paper may contribute to change this state of affairs by providing researchers with a simple score test to check the validity of the distributional assumptions in these models. If the null hypothesis is found to be too restrictive, the researcher may want to estimate a more flexible model. The simple generalisations proposed in Section 2 are attractive because they allow considerable flexibility without substantially increasing the computational cost.

#### REFERENCES

- Abramowitz, M. and Stegun, I.A., eds. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover Publications.

- Beran, R. (1988). "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687-697.
- Box, G.E.P. and Cox, D.R. (1964). "An Analysis of Transformations," *Journal of the Royal Statistical Society B*, 26, 211-243.
- Burbidge, J.B., Magee, L. and Robb, A.L. (1988). "Alternative Transformations to Handle Extreme Values of the Dependent Variable," *Journal of the American Statistical Association*, 83, 123-127.
- Cameron, A.C. and P.K. Trivedi (1998). *Regression Analysis of Count Data*, Cambridge, MA: Cambridge University Press.
- Cameron, A.C., Trivedi, P.K., Milne, F. and Piggott, J. (1988). "A Microeconomic Model of the Demand for Health Care and Health Insurance in Australia," *Review of Economic Studies*, 55, 85-106.
- Cramer, J.S. and Ridder, G. (1988). "The Logit Model in Economics," *Statistica Neerlandica*, 42, 297-314.
- Davidson, R. and MacKinnon, J.G. (1998). "Graphical Methods for Investigating the Size and Power of Hypothesis tests," *The Manchester School*, 66, 1-26.
- Hinde, J. (1982). "Compound Poisson regression models". In Gilchrist, R. (Ed.), *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, New-York: Springer-Verlag, 109-121.
- Horowitz, J.L. (1994), "Bootstrap-Based Critical Values for the Information Matrix Test," *Journal of Econometrics*, 61, 395-411.
- Horowitz, J.L. (2001) "The Bootstrap." In: Heckman, J.J., and Leamer, E. (Eds.), *Handbook of Econometrics*, Volume 5, Amsterdam: Elsevier Science, 3159-3228.
- Horowitz, J.L. and Savin, N.E. (2000). "Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach," *Journal of Econometrics*, 95, 375-389
- Judd, K. (1998). *Numerical Methods in Economics*, Cambridge, MA: MIT Press.

- MacKinnon, J.G. and Magee, L. (1990). "Transforming the Dependent Variable in Regression Models," *International Economic Review*, 31, 315-339.
- McFadden, D. and Train, K. (2000). "Mixed MNL Models of Discrete Response," *Journal of Applied Econometrics*, 15, 447-470.
- Ramsey, J.B. (1969). "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society B*, 32, 350-371.
- Winkelmann, R. (2008). *Econometric Analysis of Count Data*, 5th ed. Berlin: Springer-Verlag.
- Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.