

# Goodness-of-fit techniques for count data models: an application to the demand for dental care in Spain\*



Begoña Álvarez<sup>1</sup>, Miguel A. Delgado<sup>2</sup>

<sup>1</sup>Departamento de Economía Aplicada, Universidad de Vigo, Lagoas Marcosende, Vigo 36310, Spain.

<sup>2</sup>Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, C./ Madrid 126-128, Getafe 28903, Spain. (E-mail: delgado@est-econ.uc3m.es.)

First version received: April 2000/Final version received: March 2001

**Abstract.** This paper compares the practical performance of alternative goodness-of-fit techniques for count data models in the context of a study of the determinants of demand for dental care in Spain. We apply alternative goodness-of-fit techniques to different specifications. In particular, we implement recently proposed specification tests which are consistent in the direction of general nonparametric alternatives. The analysis suggests that a negative binomial model is an appropriate specification for dental care demand. Dental health and income are identified as important predictors of individuals' behavior.

**Key words:** Poisson regression, negative binomial regression, hurdle model, specification tests, dental care

## 1. Introduction

The application of count data models has been prominent in the analysis of the demand for health care (see, for instance, Cameron et al. 1986, Pohlmeier and Ulrich, 1994, Deb and Trivedi, 1997, Geil et al., 1997). Health economists are usually concerned with analyzing the impact of factors such as insurance, income, prices or gender on the level of health care utilization. In this case, it is well known that consistent estimates of regressor coefficients and valid inferences may be obtained applying nonlinear regression techniques,

---

\* Research funded by the Spanish "Dirección General de Enseñanza Superior", reference number PB98-0025. We are grateful to Daniel Miles and two anonymous referees for their valuable comments.

as long as the conditional mean is correctly specified (Gourieroux et al. 1984). Nonetheless, these procedures are unable to estimate the probability of single outcomes, which constitutes an important information for health policy purposes.

Consistent estimators of conditional probabilities can be obtained without imposing a parametric model, using nonparametric smoothers (see e.g. Härdle 1990 and Wand and Jones 1995). However, a full parametric specification has many advantages for the empirical researcher. On one hand, when we have several explanatory variables, the nonparametric estimates are very imprecise, because of the “curse of dimensionality” and they are extremely difficult to interpret using graphical procedures or other means. On the other hand, parameters usually have a precise interpretation, which is useful for structural analysis, policy evaluation and prediction purposes.

The reliability of any parametric specification for fitting the empirical data needs to be evaluated. Goodness-of-fit techniques are crucial for this purpose. The aim of this paper consists of comparing the practical performance of alternative goodness-of-fit techniques for count data models in the context of a study of the determinants of demand for dental care in Spain.

The data of this study are from the 1993 Spanish Health Survey. The Spanish public health system is only responsible for prevention and treatment of some acute odontological problems. So, the vast majority of dental care utilization is afforded through out-of-pocket payments by patients. In 1996, private dental expenditure accounted for 25 percent of Spanish private health expenditure and 5 per cent of total health expenditure. Moreover, the population to dentist ratio in Spain is the highest in Europe (Mossialos and Le Grand, 1999). In such a context of private provision of dental care and relative scarcity of dentists, income and prices emerge as potential predictors of dental care utilization. Since medical care has almost no substitute in the individual production of dental health, it seems important to provide empirical evidence on the existence of social disparities in the delivery of dental care derived from the effect of variables related to individuals’ ability to pay for services.

As in most empirical studies on this issue, we employ three alternative specifications for the demand for dental care: Poisson, negative binomial and hurdle negative binomial. In order to choose between the alternative specifications, we apply traditional model selection procedures consisting in comparing estimates of marginal probabilities based on empirical frequency and estimated models (Deb and Trivedi, 1997), graphical comparisons of parametric and nonparametric alternatives (see Gerfin, 1996, for an application to binary response models), and formal model selection criteria, such as  $R^2$  (Cameron and Windmeijer, 1996) and Akaike’s. We also implement new proposed formal specification tests for the functional form of the conditional mean and the conditional probability of counts, which are consistent against nonparametric alternatives (Stute, 1997 and Andrews, 1997). The test statistics are based on functionals of marked empirical processes which are not asymptotically distribution free under the null hypothesis of correct specification, but their critical values are accurately approximated by bootstrap. These formal tests are still not standard practice in empirical work, therefore our contribution aims at illustrating their performance in an application to count data.

We have organized the rest of the paper as follows. The alternative model specifications are described in section 2. In section 3 we discuss formal specification tests for the regression model and for the conditional distribution func-

tion. Spanish data used in the application are presented in section 4. Finally, the estimation results and the implementation of tests are explained in section 5. The paper concludes with some final remarks.

## 2. Count data models for dental care

Since we are concerned with cross-sectional data, observations  $\{(Y_i, X_i), i = 1, \dots, n\}$  are assumed to be independent and identically distributed as a random vector  $(Y, X)$ ; where  $Y$  are count data variables denoting the number of visits to the dentist, and  $X = (X_1, X_2, \dots, X_k)'$  is the vector of explanatory variables.

We consider two alternative approaches to dental care utilization. First, we consider models where utilization is decided in one stage, as suggested by Cameron et al. (1988), among others. In this case, the conditional probabilities of counts are represented by

$$P_y(x) = \Pr(Y = y | X = x).$$

Many alternative parametric specifications have been proposed in the literature, the most popular being the Poisson and negative binomial, which are given by

$$P_y(x) = \frac{\exp(-m(x))m(x)^y}{y!} \quad (1)$$

and

$$P_y(x) = \frac{\Gamma(y + \delta_0^{-1})}{\Gamma(\delta_0^{-1})\Gamma(y + 1)} \left( \frac{\delta_0^{-1}}{m(x) + \delta_0^{-1}} \right)^{\delta_0^{-1}} \left( \frac{m(x)}{m(x) + \delta_0^{-1}} \right)^y \quad (2)$$

respectively, where

$$m(x) = E(Y | X = x)$$

and  $\delta_0 > 0$  is a parameter, which allows more dispersion in the negative binomial specification with respect to the Poisson model.

As in most previous studies (see Cameron and Trivedi's, 1998 book for a fairly comprehensive survey of empirical work using these models), we assume that

$$m(x) = \exp(\alpha_0 + x'\beta_0).$$

The exponential conditional mean permits us to interpret coefficients as the proportional change in  $m(x)$ , when the corresponding explanatory variable changes by one unit. If the explanatory variable enters logarithmically in the regression model, then its coefficient is an elasticity. Finally, if the  $k$ -th explanatory variable is an indicator, then the conditional mean is  $\exp(\beta_k)$  times

larger if the indicator variable is unity rather than zero (see Cameron and Trivedi, 1998 for analytical details).

Indeed, consistent estimates of the parameters in the conditional mean are possible by applying nonlinear least squares, or generalized least squares in order to improve efficiency (see Delgado and Kniesner 1997 for discussion).

The second approach consists of a two-part model. The first part specifies the contact decision with dental care services, whilst the second part specifies the frequency of visits, once contact has been established. This two-part specification is sensible in a dental care context. In fact, there are reasons to believe that explanatory variables perform different roles depending on the stage of the decision-making process. If we consider a dental treatment spell, the first consultation usually consists of an examination – sometimes free of charge – or a preventive control visit. In that first visit, individuals obtain information on their “care needs” – patients’ needs as perceived by the dentist – and, usually, on the expected duration of treatment. Thus, it seems reasonable to assume that a contact decision is more responsive to attitudes toward dental health, whilst financial constraints become more relevant at the second stage of decision. Furthermore, as private dentists are remunerated on a fee-for-service basis, they face incentives to induce demand at that second decision.

Following Pohlmeier and Ulrich (1995) and Gurmu and Trivedi (1996), we model the two-part decision by the hurdle version of the negative binomial specification. Under this specification, contact probability is governed by the following process,

$$P(Y = 0 | X) = \left( \frac{v_1}{v_1 + \lambda_1(X)} \right)^{v_1}. \quad (3)$$

Note that when  $v_1 = 1$  the binary process is a logit model.

When an individual crosses the hurdle, that is, when a contact with dental services has been established, frequency decision is conditionally distributed as a truncated negative binomial,

$$P(Y = y | y > 0, X) = \frac{\Gamma(y + v_2)}{\Gamma(y + 1)\Gamma(v_2)} \left[ \left( \frac{v_2 + \lambda_2(X)}{v_2} \right)^{v_2} - 1 \right]^{-1} \left( \frac{\lambda_2(X)}{v_2 + \lambda_2(X)} \right)^y, \quad (4)$$

where  $\lambda_j(x)$  is a function of  $x$  and  $v_j = \delta_{0j}^{-1}$ ,  $j = 1, 2$ . As in previous studies, we consider the parameterization  $\lambda_j(x) = \exp(\alpha_{0j} + x'\beta_{0j})$ , where  $(\alpha_{0j}, \beta_{0j})$ ,  $j = 1, 2$ , represent unknown parameters. Therefore, the conditional probability of counts can be written as

$$P_y(X) = [P(Y = 0 | X)]^d [(1 - P(Y = 0 | X)) \times P(Y = y | y > 0, X)]^{1-d} \quad (5)$$

where  $d = 1$  if  $y = 0$  and  $d = 0$  otherwise,  $P(Y = 0 | X)$  and  $P(Y = y | y > 0, X)$  are given in (3) and (4) respectively.

Under this specification, the conditional mean of the count variable is given by

$$E(Y|X) = \left[ 1 - \left( \frac{v_1}{v_1 + \lambda_1(X)} \right)^{v_1} \right] E(Y | Y > 0, X),$$

where

$$E(Y | Y > 0, X) = \frac{\lambda_2(X)}{\left[1 - \left(\frac{v_2}{v_2 + \lambda_2(X)}\right)^{v_2}\right]}.$$

It is important to remark that the hurdle negative binomial and the standard negative binomial are identical when  $\delta_{01} = \delta_{02}$  and  $\lambda_1(X) = \lambda_2(X)$ .

The two parts of the hurdle model are assumed to be functionally independent, therefore the joint likelihood may be maximized by separately maximizing the log-likelihoods of the two components: the right censored-at-one negative binomial and the left truncated-at-zero negative binomial.

### 3. Goodness-of-fit tests

In this section, Stute (1997) and Andrews (1997) tests are proposed as formal specification tests for the functional form of the conditional expectation and the conditional probabilities of counts, respectively. These tests are consistent in the direction of general nonparametric alternatives. The corresponding test statistics are based on marked empirical processes which are asymptotically not distribution free, but their critical values are well approximated by bootstrap, as we will describe below.

#### 3.2. Goodness-of-fit tests for regression and binary decision models

##### 3.2.1. *A counting decision process.* —

Research on count data has placed considerable emphasis in the correct specification of the conditional mean. Consistent estimators of the conditional mean parameters are available without full information on the whole underlying data generation process (see e.g. Gourieroux et al., 1984, Delgado and Kniesner, 1997 or Cameron and Johansson, 1997, among others). Furthermore, some common specification tests, such as overdispersion tests, are also based on the correct specification of the conditional mean (Cameron and Trivedi, 1990). Despite the relevance of this assumption, applied count data literature has been silent with regard its misspecification.

In this subsection we present a formal test for the null hypothesis

$$H_0 : E(Y|X) = m(X; \beta_0) \quad \text{a.s. for some } \beta_0 \in \mathbf{B}, \quad (6)$$

consistent in the direction of general nonparametric alternatives, which are the negation of the null. In the previous expression  $m$  is a known function,  $\beta_0 \in \mathbf{B}$  is an unknown parameter vector and  $\mathbf{B}$  is the parameter space, a compact subset of  $\mathbb{R}^k$ .

Formal specification tests for regression curves based on smoothers are available in generous supply (see e.g. Härdle and Mammen, 1993, Eubank and Spiegelman, 1990, Hong and White, 1995, Zheng, 1998, to mention only a few). These tests have the disadvantage of depending on a smoothing number, the

bandwidth, and so far no rules are available for an automatic choice of such number.

Here, we apply a test based on the fact that  $H_0$  can be written as,

$$H_0 : T(\bar{X}, \beta_0) = 0, \quad a.s. \text{ for some } \beta_0 \in \mathbf{B} \quad (7)$$

where

$$\begin{aligned} T(x, \beta_0) &= \int_{-\infty}^x (E(Y | X = u) - m(u; \beta_0)) dF_x(u) \\ &= E\{[Y - m(X; \beta_0)]1(X \leq x)\} \end{aligned}$$

and  $F_x(\cdot)$  is the distribution function of  $X$  and  $1(A)$  is the indicator function of the event  $A$ . Note that, with  $X_i$  multivariate we use the notation,

$$1(X_i \leq x) = \prod_{j=1}^k 1(X_{ji} \leq x_j).$$

That is,  $T$  is the difference between the integrated regression curves under the specifications in the parametric null and in the nonparametric alternative.

A natural consistent estimator of  $T(x, \beta_0)$  under  $H_0$  is,

$$\hat{T}_n(x) = \frac{1}{n} \sum_{i=1}^n (Y_i - m(X_i; \hat{\beta}_n))1(X_i \leq x),$$

where  $\hat{\beta}_n$  is some  $\sqrt{n}$ -consistent estimator. The test is based on some functional of  $n^{1/2}\hat{T}_n$ , for instance, the Cramér-von Mises statistic

$$\hat{C}_n = \frac{1}{n} \sum_{i=1}^n [n^{1/2}\hat{T}_n(X_i)]^2.$$

Stute (1997) has obtained the limiting process of  $n^{1/2}\hat{T}_n$  for general nonlinear regression models, under very weak regularity conditions, which depends on certain unknown features of the underlying data generating process. Therefore, asymptotic tests are difficult to implement, except in particular circumstances. Stute et al. (1998) suggested the application of a residual based ‘wild bootstrap’ for estimating the limiting distribution of  $\hat{C}_n$ . The ‘wild bootstrap’ was introduced by Wu (1986) in order to deal with conditional heteroskedastic disturbances in regression models.

The bootstrap resample procedure is as follows:

1. Obtain a  $\sqrt{n}$ -consistent estimator under  $H_0$ ,  $\hat{\beta}_n$  say, for the original sample  $\mathcal{X}_n = \{(Y_i, X_i), i = 1, \dots, n\}$  and compute the statistic  $\hat{C}_n$ .
2. Generate i.i.d.  $V_i, i = 1, \dots, n$  independent of  $X_i$ , bounded, and such that  $E(V_i) = 0, E(V_i^2) = 1$ . In practice, it is recommended that these variables match the conditional skewness properties of the underlying regression errors, based on second order asymptotic considerations (see Liu, 1988). To

this end,  $E(V_i^3) = 1$  is usually chosen in practice. In our applications, we have selected a distribution for  $V_i$  discrete, taking values  $-(\sqrt{5} + 1)/2$  and  $(\sqrt{5} + 1)/2$ , with attached masses,  $(\sqrt{5} + 1)/2\sqrt{5}$  and  $(\sqrt{5} - 1)/2\sqrt{5}$ , respectively. This distribution for  $V_i$  was suggested by Liu (1988), and it has been recommended in specification-testing problems by Härdle and Mammen (1993) and Stute et al. (1998), among others. Simulation results provided by Stute et al. (1998) are most encouraging.

3. Construct a new resample  $\mathcal{X}_n^* = \{(Y_i^*, X_i), i = 1, \dots, n\}$  from the original sample  $\mathcal{X}_n$  where  $Y_i^*$  is generated as

$$Y_i^* = m(X_i; \hat{\beta}_n) + \varepsilon_i^*,$$

with  $\varepsilon_i^* = \hat{\varepsilon}_i V_i$ , and  $\hat{\varepsilon}_i = Y_i - m(X_i; \hat{\beta}_n)$ .

4. With the new bootstrap resample  $\mathcal{X}_n^*$ , compute the bootstrap analogs of  $\hat{\beta}_n$ ,  $\hat{\beta}_n^*$  say, and bootstrap versions of  $\hat{C}_n$ ,

$$\hat{C}_n^* = \frac{1}{n} \sum_{i=1}^n [n^{1/2} \hat{T}_n^*(X_i)]^2,$$

where,

$$\hat{T}_n^*(x) = \frac{1}{n} \sum_{i=1}^n (Y_i^* - m(X_i, \hat{\beta}_n^*)) 1(X_i \leq x).$$

5. Repeat steps 2 to 4  $B$  times, yielding resamples  $\mathcal{X}_n^{*(b)}$ ,  $b = 1, \dots, B$ , and, for each resample, compute the bootstrap version of  $\hat{C}_n$ ,  $\hat{C}_n^{*(b)}$ ,  $b = 1, \dots, B$ , with  $B$  large, as large as required for accuracy.

The critical values of the test are estimated by the conditional quantiles, given the sample  $\mathcal{X}_n$ , of  $\hat{C}_n^*$ , which are approximated by the quantiles of  $\hat{C}_n^{*(b)}$ ,  $b = 1, \dots, B$ . That is, at the  $\alpha$ -level of significance, the bootstrap critical value is  $\hat{C}_{\alpha n}^*$ , given by,

$$\Pr(\hat{C}_n^* > \hat{C}_{\alpha n}^* | \mathcal{X}_n) = \alpha,$$

which is difficult to compute exactly in practice, but it can be approximated by  $\hat{C}_{\alpha n}^{*(B)}$ , where,

$$\frac{1}{B} \sum_{b=1}^B 1(\hat{C}_n^{*(b)} > \hat{C}_{\alpha n}^{*(B)}) = \alpha.$$

The bootstrap  $p$ -values are approximated by

$$p_B^* \text{-value} = \frac{1}{B} \sum_{b=1}^B 1(\hat{C}_n^{*(b)} > \hat{C}_n).$$

The accuracy of such approximations depends on how large  $B$  is. So, at the  $\alpha$ -level of significance,  $H_0$  is rejected when  $\hat{C}_n > \hat{C}_{\alpha n}^{*(B)}$  or  $p_B^* \text{-value} \leq \alpha$ . Critical

values and  $p$ -values are computed in the same way for all the tests described below.

Stute et al. (1998) showed that this test is consistent in the sense that the conditional distribution of the bootstrap test statistic, given the sample, consistently estimates the asymptotic distribution of the test statistic under  $H_0$ . Under the alternative hypothesis,  $H_1$ , the conditional distribution – given the sample – of the bootstrap test statistic converges in probability to a proper distribution, and bootstrap critical values are bounded in probability. Since under  $H_1$  the test statistic diverges to infinity, it will be greater than the bootstrap critical value with probability one, and the probability of a type two error converges to zero.

### 3.2.2. *A binary decision process.*—

Researchers may eventually be only concerned with the binary decision between zero counts and positive counts. In a context of health care demand, this decision is interesting in itself because it is very sensitive to determinants of access to health services, which constitutes relevant information for health policy purposes. In the hurdle model, the decision to seek care governs the first part of the specification. In fact, if the number of users is extremely high and the positive outcomes have short tails, the explanatory power of predictors is likely to be concentrated in this part of the model (Gurmu, 1997).

Mullahy (1997) develops a class of goodness-of-fit tests to detect mixture departures from the maintained null hypothesis of a probability model (taken in his paper to be the Poisson model). Those tests are also based on comparison of observed zero outcomes and the proportion expected in the zero cell under the null model.

Noting that the conditional probability of zero counts can be expressed as a conditional expectation, that is

$$P_0(x) = \Pr(Y = 0 | X = x) = E[1(Y = 0) | X = x],$$

it is fairly easy to extend the ideas for specification testing of a regression curve to the specification testing of a conditional distribution of a binary decision process. Thus, the null hypothesis

$$H_0 : P_0(X) = P_0(X; \theta_0) \quad \text{a.s. for some } \theta_0 \in \Theta,$$

where  $P_0$  follows some particular parametric distribution evaluated at zero, is of the same form as (6). Also  $H_0$  can be written as (7)

$$H_0 : M(0, X; \theta_0) = 0 \quad \text{a.s. for some } \theta_0 \in \Theta$$

where

$$\begin{aligned} M(0, x; \theta_0) &= \int_{-\infty}^x \{E[1(Y = 0) | X = u] - P_0(u; \theta_0)\} dF_x(u) \\ &= E\{[1(Y = 0) - P_0(X; \theta_0)]1(X \leq x)\}. \end{aligned}$$

Following Stute (1997) and Andrews' (1997) proposals, we can develop a test for this hypothesis based on functionals of the estimated integrated regression function

$$\hat{M}_n(0, x) = \frac{1}{n} \sum_{i=1}^n \{1(Y_i = 0) - P_0(X_i; \hat{\theta}_n)\} 1(X_i \leq x).$$

Then, a Cramér-von Mises statistic is given by

$$\tilde{C}_n = \frac{1}{n} \sum_{i=1}^n [n^{1/2} \hat{M}_n(0, X_i)]^2.$$

The statistic  $\tilde{C}_n$  is not distribution free and a bootstrap test is also well motivated. Since full information on the data generating process is available under the null, a parametric bootstrap method, suggested by Andrews (1997) in this context, can be implemented as follows:

1. Obtain the maximum likelihood estimator under  $H_0$ , say  $\hat{\theta}_n$ , for the original sample  $\mathcal{X}_n = \{(Y_i, X_i), i = 1, \dots, n\}$  and compute the statistic  $\tilde{C}_n$ .
2. Construct a new resample  $\mathcal{X}_n^* = \{(Y_i^*, X_i), i = 1, \dots, n\}$  from the original sample  $\mathcal{X}_n$ , where  $Y_i^*$  is generated as a Bernoulli random variable with parameter  $P_0(X_i; \hat{\theta}_n)$ ,  $P_0$  in the null, for each  $X_i$ .
3. With the new bootstrap sample, compute the bootstrap analogs of  $\hat{\theta}_n$ ,  $\hat{\theta}_n^*$  say, and the bootstrap version of  $\tilde{C}_n$ ,

$$\tilde{C}_n^* = \frac{1}{n} \sum_{i=1}^n [n^{1/2} \hat{M}_n^*(0, X_i)]^2,$$

where

$$\hat{M}_n^*(0, x) = \frac{1}{n} \sum_{i=1}^n \{1(Y_i^* = 0) - P_0(X_i; \hat{\theta}_n^*)\} 1(X_i \leq x).$$

4. Repeat steps 2 and 3  $B$  times, yielding resamples  $\mathcal{X}_n^{*(b)}$ ,  $b = 1, \dots, B$ , and, for each resample, compute the bootstrap versions of  $\tilde{C}_n$ ,  $\tilde{C}_n^{*(b)}$ ,  $b = 1, \dots, B$ , with  $B$  large.

The critical values and  $p$ -values are approximated from  $\tilde{C}_n^{*(b)}$ ,  $b = 1, \dots, B$ , as in the previous test.

### 3.3. A goodness-of-fit test for the conditional distribution function

Despite, in most applied work on count data, the interest being focused on estimating the conditional mean and its associated partial effects, in many cases, analysts might wish to use the estimates for predictive purposes or simply to learn the underlying data generating process (Winkelmann and Zimmerman, 1995). In this case, consistency of coefficient estimates and validity of inferences require the correct specification of the conditional distribution of counts.

Formally, the hypothesis we want to test is

$$H_0 : P_y(X) = P_y(X; \theta_0) \quad \text{a.s. for some } \theta_0 \in \Theta \text{ and all } y = 0, 1, 2, \dots \quad (8)$$

where  $P_y$  is a parametric distribution function and  $\theta_0$  is an unknown parameter vector. When the alternative hypothesis,  $H_1$ , is parametric, traditional tests can be applied. If  $H_1$  nests  $H_0$ , LM, LR or Wald tests can be applied, and if  $H_1$  is non-nested we can apply the LR test of Cox (1962).

For instance, model simplification going from the negative binomial to the Poisson model and from the hurdle negative binomial to the standard negative binomial involves nested hypotheses. Since the negative binomial model collapses to the Poisson distribution when  $\delta_0 \rightarrow 0$ , a specification test for the Poisson model, which is consistent in the direction of the negative binomial, can be implemented by testing the parametric null hypothesis  $H_0 : \delta_0 = 0$  against  $H_1 : \delta_0 > 0$ . In this case, the distribution of LR and Wald test statistics are non-standard because the null hypothesis lies on the boundary of the parameter space for the negative binomial, which does not permit underdispersion. Nonetheless, one can still implement regression-based score tests of the null hypothesis of zero overdispersion (Cameron and Trivedi, 1990). Model simplification going from the hurdle negative binomial to the standard negative binomial involves the nested hypothesis,  $H_0 : (\beta_{01}, \delta_{01})' = (\beta_{02}, \delta_{02})'$ , which could be tested for by implementing LR and Wald tests.

Once a model has been selected against the competing alternatives, goodness-of-fit tests should be used to evaluate whether it provides a good fit of the data. For this purpose, let us re-express the null hypothesis (8) as

$$H_0 : M(y, X; \theta_0) = 0 \quad \text{a.s. some } \theta_0 \in \Theta \text{ and } y = 0, 1, 2, \dots$$

where

$$\begin{aligned} M(y, x; \theta_0) &= \int_{-\infty}^x \{E(Y = y | X = u) - P_y(u; \theta_0)\} dF_x(u) \\ &= E\{[1(Y = y) - P_y(X; \theta_0)]1(X \leq x)\}. \end{aligned}$$

The alternative, as before, is the negation of the null, that is, any other data-generating process.

The function  $M(y, x; \theta_0)$  can be estimated by

$$\hat{M}_n(y, x) = \frac{1}{n} \sum_{i=1}^n \{1(Y_i = y) - P_y(X_i; \hat{\theta}_n)\} 1(X_i \leq x),$$

where  $\hat{\theta}_n$  is the maximum likelihood estimate of  $\theta_0$ . Andrews (1997) have justified tests of  $H_0$  based on functionals of  $\hat{M}_n(y, x)$ . He studies the Kolmogorov-Smirnov type

$$\bar{K}_n = \max_{1 \leq i \leq n} |n^{1/2} \hat{M}_n(Y_i, X_i)|,$$

but his theory is also valid for the Cramér-von Mises statistic

$$\bar{C}_n = \frac{1}{n} \sum_{i=1}^n [n^{1/2} \hat{M}_n(Y_i, X_i)]^2$$

which can also be implemented. Asymptotic tests are not available either in this case because the limiting distribution of the statistics is case dependent. Andrews (1997) has justified the consistency of parametric bootstrap tests in this context. The parametric bootstrap is performed as follows.

1. Obtain the maximum likelihood estimator under  $H_0$ , say  $\hat{\theta}_n$ , for the original sample  $\mathcal{X}_n = \{(Y_i, X_i), i = 1, \dots, n\}$  and compute the statistic  $\bar{C}_n$ .
2. Construct a new resample  $\mathcal{X}_n^* = \{(Y_i^*, X_i), i = 1, \dots, n\}$  from the original sample  $\mathcal{X}_n$ , where  $Y_i^*$  is generated as a discrete random variable with distribution  $P_y(X_i; \hat{\theta}_n)$ ,  $P_y$  in the null, for each  $X_i$ .
3. With the new bootstrap sample, compute the bootstrap analogs of  $\hat{\theta}_n$ ,  $\hat{\theta}_n^*$  say, and the bootstrap version of  $\bar{C}_n$ ,

$$\bar{C}_n^* = \frac{1}{n} \sum_{i=1}^n [n^{1/2} \hat{M}_n^*(Y_i^*, X_i)]^2,$$

where,

$$\hat{M}_n^*(y, x) = \frac{1}{n} \sum_{i=1}^n \{1(Y_i^* = y) - P_y(X_i; \hat{\theta}_n^*)\} 1(X_i \leq x).$$

4. Repeat steps 2 and 3  $B$  times, yielding resamples  $\mathcal{X}_n^{*(b)}$ ,  $b = 1, \dots, B$ , and, for each resample, compute the bootstrap versions of  $\bar{C}_n$ ,  $\bar{C}_n^{*(b)}$ ,  $b = 1, \dots, B$ , with  $B$  large.

The critical values and  $p$ -values are approximated from  $\bar{C}_n^{*(b)}$ ,  $b = 1, \dots, B$ , as in previous tests.

#### 4. Data on dental care demand

The following application uses data from the 1993 Spanish Health Survey (SHS). The total sample size of the survey is 21,063 people. In this study, we have restricted ourselves to the 6,258 working individuals, aged 16 to 65 who answered all relevant questions. The survey offers information concerning health-related issues, as well as demographic and socioeconomic characteristics of respondents. In order to control for heterogeneity, we have divided our sample into two subsamples according to gender of respondents. The existence of gender difference in health care utilization is well documented (e.g. Sindelar, 1982 and Hunt-McCool et al. 1994). Different use of dental care by gender may arise because of underlying biological conditions and general attitudes towards dental health.

The dependent variable in the model is the number of visits to the dentist over the three months before the interview. Table 1 reports actual frequency distribution of visits to the dentist in both subsamples. The zero counts account for about 80 percent in the sample of working women and 86 percent in the

**Table 1.** Frequency of visits to the dentist

	0	1	2	3	4	5	6	7	8+	Total
Working women	1442	178	75	39	20	8	10	3	16	1791
Working men	3835	325	142	69	46	23	11	5	11	4467

**Table 2.** Type of financing of the last visit

	Public insurance	Private insurance	Out-of-pocket
Working women	5.7	5.1	88.5
Working men	8.5	2.7	88.1

male sample. Right tails are not very long. Around 0.9 percent of the female patients and 0.2 percent of male patients have 8 or more visits.

The theoretical framework on the demand for medical care applies directly to dental care. Following Cameron et al. (1988), we assume that the demand for health care depends on health status and other demographic variables, such as age (*AGE*) and marital status (*MARRIED*). Since dental health status cannot be measured directly with SHS data, dichotomous variables based on the number of teeth of respondents are used as dental health proxies. There are three categories for describing dental health status: individuals conserve all their own teeth (*DENTH1*), more than a half of their own teeth (*DENTH2*) or less than a half of their own teeth (*DENTH3*). These denture conditions are shown to be an important predictor of dental care utilization in other studies (Grytten et al. 1990).

There are other factors, besides the biological ones, that determine the demand for dental care. Economic studies have focused mainly on health insurance, income, prices, and education. As we pointed out before, the Spanish National Health Service only covers a small package of dental services. Therefore, the vast majority of Spaniards, about 88 percent in our samples (see Table 2), afford dental care through private payments.

This institutional setting is likely to generate moral hazard incentives which may induce individuals' demand for private insurance in response to expected dental care needs. However, the SHS suggests that such a coverage is quite moderate: only 5.1 percent of women and 2.7 percent of men had the last visit covered by a private insurance. This may be due to the high co-payments that private insurance companies establish for dental services, in comparison with those corresponding to other medical services. Information on dental insurance coverage is only available from those respondents who used dental care during the reference period, so we cannot include it as an explanatory variable in our models.

As out-of-pocket payments represent a high share of health care financing, ability to pay becomes an important determinant of health care use. Then, a positive effect of household income level on the demand for dental care is expected. Since the SHS does not provide information on this variable, income values are recovered from an income equation estimated with data from the 1990–1991 Spanish Household Expenditure Survey (SFES). The SFES contains

**Table 3.** Definition of variables

NAME	DEFINITION
DENTIST VISITS	Number of visits to the dentist in the 3 months previous to the interview
AGE	Age of respondent $\times 10^{-1}$
DENTH1	Dichotomous variable for dental health (= 1 if respondent has all her/his natural teeth, = 0 otherwise)
DENTH2	Dichotomous variable for dental health (= 1 if respondent has more than half of her/his own teeth, = 0 otherwise)
DENTH3	Dichotomous variable for dental health (= 1 if respondent has less than half of her/his own teeth, = 0 otherwise)
MARRIED	Dichotomous variable (= 1 if married, = 0 otherwise)
Log(INCOME)(*)	Estimated logarithm of household income
SCHOOL	Years of schooling completed
DENTISTS	Number of dentists per 100.000 population at provincial level. There are 52 provinces. Source: INE
PRICE INDEX	Medical price index component of Consumption Price Index at provincial level. Source: INE

(\*) To create this variable, we have estimated a linear regression model with data from the 1990–91 Spanish Household Expenditures Survey. Dependent variable was total expenditure of households and as explanatory variables we considered: average age of adults, number of adults, number of children, education level of head of household, labor status of head of household, size of town and region of residence.

comprehensive information on income, expenditure and socio-demographic characteristics of 21,115 households and their members. The income values (*INCOME*) are estimated by their linear predictors, given a set of explanatory variables common in both data sets (see Table 3, footnote). A potential advantage of this approach is that it overcomes the usual problems generated by under-reported income information in interviews. Details on the complementary use of both surveys are described in Álvarez (2001).

Education is often found to be relevant in explaining the demand for health care. However, its effect is complex. It has been argued that more educated people are more prone to follow healthy life-styles. For example, previous studies have found that the demand for preventive medical services increases with education (Kenkel, 1994). Grossman (1972) suggests that schooling may improve the efficiency of individuals health production which would reduce their health care utilization. An alternative explanation is provided by Fuchs (1982), who considers that time preference causes investment in both health and education. This means that the observed correlation between health care demand and education does not necessarily imply a causal relationship. In order to test differences in Spanish dental care utilization, according to education, we include the variable years of formal education (*SCHOOL*).

Availability of dental services is measured by means of a dentist per population ratio at provincial level (*DENTIST*). This variable accounts for two facts. First, if differential of mobility between consumers and producers of dental care generates local variations in excess demand for dental care, the number of dentists per population performs as a non-price rationing device (Manning and Phelps, 1979). That is, the lower the dentist per population ratio, the lower the availability of dental services and the higher the non-monetary

opportunity costs of using them (e.g. time costs). Second, this variable can also proxy the supplier-induced response of demand. When the supply of dental services is high, dentists could compensate the expected earnings fall by stimulating demand without lowering fees. These two effects tend to increase demand for dental care. Consequently, a positive sign for dentists density is expected.

Finally, dental care prices are proxied by the index component of outpatient and nursery services in the Consumption Price Index (*PRICES*) at provincial level. It is clear that such index constitutes an imperfect measure of dental care prices. On the one hand, it aggregates prices for all outpatient private services, including dental services. On the other hand, since Spanish dentists' fees are set freely, to consider a broad geographical index ignores the possible dispersion of prices within any province. Despite these criticisms, we include this variable because otherwise it would be difficult to determine whether increased utilization with increased supply of dentists is due to inducement or to falling prices.

Definition of variables used in this study is reported in Table 3 and summary statistics are presented in the first columns of Tables 4 and 5.

## 5. Estimation results

The point estimates of models presented in section 2 are reported in Tables 4 and 5. Columns 2 to 5 display maximum likelihood estimates from the standard Poisson and negative binomial specifications, with  $m(x) = \exp(\alpha_0 + x'\beta_0)$ . In both models, the magnitude of estimates are more or less the same though, in some cases, significance levels differ markedly.

Estimates obtained from the hurdle specification are reported in the four last columns. We assume  $\lambda_j(x) = \exp(\alpha_{0j} + x'\beta_{0j})$ ,  $j = 1, 2$ . As in most previous work, the first part of the hurdle has been specified as a logit model by setting  $\delta_{01} = 1$ , which implies a reparameterization of the model. Despite our data not offering information on treatment spells, the short observation period – three months – reduces the probability that visits correspond to multiple illness episodes. Of course, this hardly means that we observe a complete spell but, at least, misclassification of visits is not likely to be a major problem.

The inclusion of an imputed regressor measured with sampling error-estimated household income-implies that the usual standard errors are inconsistent estimators of the standard deviations. So, we computed asymptotically correct robust standard errors, following the general procedure suggested by Murphy and Topel (1985). Tables also report the  $R^2$ 's based on deviance residuals (Cameron and Windmeijer, 1996), as well as the Consistent Akaike Information Criterion (*CAIC*) defined as

$$CAIC = -2\ln \hat{l} + (1 + \ln n)p$$

where  $\hat{l}$  is the fitted likelihood,  $n$  is the sample size and  $p$  the number of parameters in the model.

In subsection 5.1, we implement Stute (1997) specification test for the functional form of the conditional expectation. In order to discriminate among competing distributional models, we apply several model selection techniques that have been used in the count data literature. Finally, we

**Table 4.** Working women ( $N = 1791$ )

	POISSON			NEG.BIN.			HURDLE NEG.BIN		
	(mean/std)	coef.	t-ratio	coef.	t-ratio	coef.	t-ratio	coef.	t-ratio
CONST.		-4.200	1.199	-2.290	0.650	-8.933	2.698	7.953	1.663
AGE	(3.53/1.14)	-0.279	3.692	-0.279	3.809	-0.227	3.243	-0.179	1.949
DENTH2	(0.42/0.49)	0.494	3.602	0.470	3.505	0.297	2.256	0.529	2.922
DENTH3	(0.09/0.28)	0.379	1.112	0.225	0.699	-0.085	-0.305	0.931	2.278
MARRIED	(0.55/0.49)	0.076	0.525	0.110	0.812	0.118	0.882	-0.063	0.341
Log(INCOME)	(14.9/0.45)	0.179	1.159	0.130	0.861	0.370	2.326	-0.335	1.664
SCHOOL	(11.1/4.16)	0.007	0.467	0.018	1.180	0.024	1.526	-0.013	0.570
DENTIST	(30.8/9.48)	0.016	2.044	0.017	2.092	0.011	1.661	0.013	1.364
PRICE INDEX	(105.2/2.7)	0.006	0.267	-0.005	0.204	0.018	0.781	-0.030	0.870
$\delta$				5.24	11.43			2.66	1.93
$R^2$		0.039		0.039					
Log-lik		-1850.21		-1393.64		-860.44			-520.34
<i>CAIC</i>		3776.8		2872.1		2922.7			

Table 5. Working men ( $N = 4467$ )

	POISSON			NEG.BIN.			HURDLE NEG.BIN				
	(mean/std)			coef.	t-ratio	coef.	t-ratio	coef.	t-ratio	coef.	t-ratio
CONST.				-5.096	1.982	-5.628	2.188	-2.780	1.130	-6.076	1.753
AGE		(3.93/1.17)		-0.068	1.414	-0.065	1.371	-0.065	1.457	-0.016	0.269
DENTH2		(0.46/0.49)		0.399	4.146	0.448	4.822	0.333	3.569	0.294	2.331
DENTH3		(0.08/0.28)		1.036	6.324	1.190	7.113	0.577	3.459	1.123	6.178
MARRIED		(0.71/0.45)		-0.148	1.297	-0.198	1.738	-0.048	0.448	0.272	1.866
Log(INCOME)		(14.9/0.43)		0.221	1.750	0.264	2.146	0.013	0.190	0.461	2.837
SCHOOL		(10.3/3.99)		0.059	4.439	0.062	4.747	0.074	6.059	0.002	0.175
DENTIST		(30.1/10.0)		0.0006	0.127	0.002	0.436	0.002	0.626	0.0006	0.093
PRICE INDEX		(105.1/2.6)		-0.0004	0.022	-0.002	0.110	-0.0002	-0.016	-0.008	0.415
$\delta$				6.931	16.094					1.11	3.334
$R^2$		0.037		0.038							
Log-lik		-3514.33		-2689.58					-1788.33		-883.259
CAIC		7113.2		5472					5521.7		

implement Andrews (1997) specification test for the functional form of the conditional probabilities of counts. Subsection 5.2 is devoted to explaining the main economic results of estimations.

### 5.1. Implementing goodness-of-fit techniques

First of all, we apply the Stute (1997) specification test for the functional form of the conditional expectation of counts. The null hypothesis is

$$H_0 : E(Y | X) = \exp(\alpha_0 + X'\beta_0) \text{ a.s.}$$

and the alternative is the negation of  $H_0$ . In order to implement this test, we generated 500 ‘wild bootstrap’ samples (see section 3.2.1), taking the Poisson estimates as the initial consistent estimates under the null. The bootstrap  $p$ -values were 0.182, for the female sample, and 0.980 for the male sample (see row 1 of Table 7). Thus, in both cases, we are unable to reject  $H_0$ , for any reasonable significance level, though the  $p$ -value is smaller for the female sample. Provided the conditional mean function is correctly specified, consistency holds for the Poisson estimates of coefficients, even if this is not the true underlying data-generating process, and valid inferences may be obtained from estimated robust standard errors.

Discriminating among the three proposed count data models on the basis of this result is, however, not possible. The null hypothesis is compatible with the Poisson, negative binomial and hurdle specifications.

If our interest is focused on the binary decision of seeking dental care, a natural hypothesis to be tested is that the conditional probability of the zero count follows a logit model. That is,

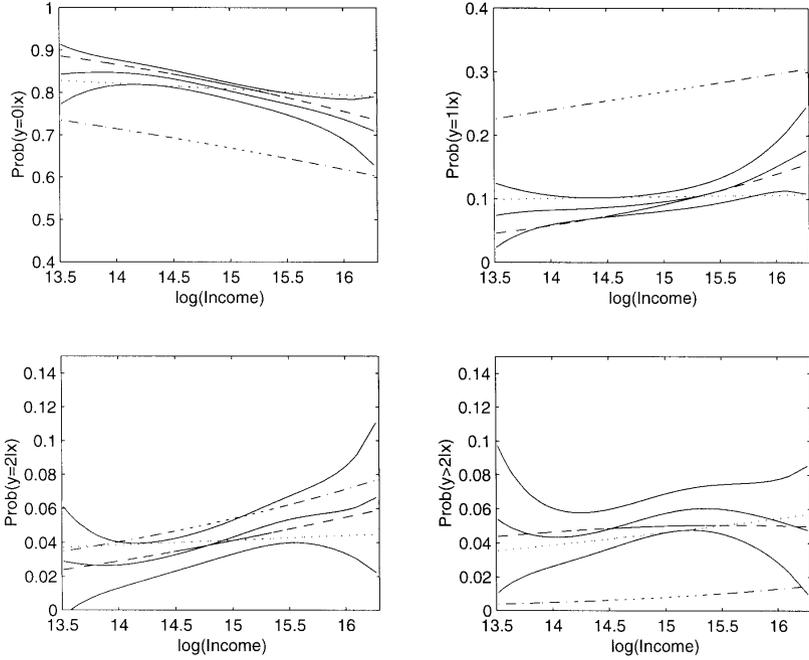
$$H_0 : E[1(Y = 0) | X] = \frac{1}{1 + \exp(\alpha_{01} + X'\beta_{01})} \text{ a.s..}$$

This specification corresponds to a negative binomial with  $\delta_0 = 1$ , as assumed in the first part of the hurdle model. We generated 500 bootstrap samples assuming a logistic data generating process with parameters given by the logit estimates shown in the sixth columns of Tables 4 and 5. The bootstrap  $p$ -values of the test are presented in row 2 of Table 7. The null hypothesis is not rejected in the two samples at usual significance levels.

To select between the alternative distributional assumptions, we first apply several selection criteria. Looking at the  $R^2$  based on deviance residuals, it is not possible to select between the Poisson and negative binomial model, since the statistics are identical for both models. As for the likelihood, the hurdle negative binomial likelihood is the greater in both subsamples. However, the  $CAIC$  is slightly smaller for the negative binomial specification than for the other two models, suggesting the selection the negative binomial under this criterion.

Graphical procedures based on nonparametric estimations are often employed in discrete data models both for exploratory analysis and for testing for the parametric specification of a given model. Figures 1 and 2 represent the parametric conditional probability estimates against income, where the other

- Nonparametric estimates and 95 % confidence level bands
- - - - Poisson
- ..... Negative Binomial
- - - - Hurdle Negative Binomial



**Fig. 1.** Estimated conditional probabilities for working women (Other variables fixed at their mean value.)

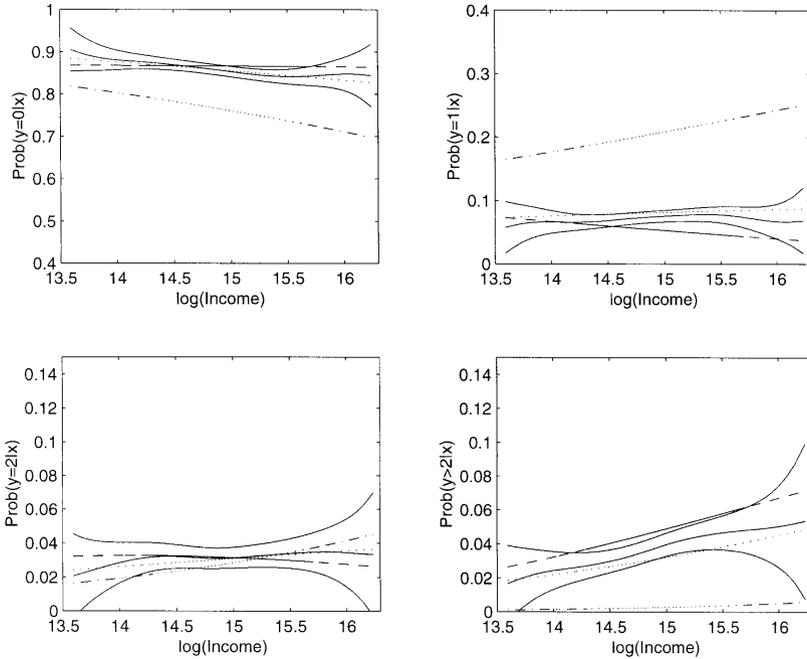
variables have been set at their mean value. Parametric estimates are compared with the nonparametric estimate (see e.g. Härdle, 1990 or Wand and Jones, 1995) computed as

$$\tilde{P}_{ny}(x) = \frac{\sum_{i=1}^n 1(Y_i = y) K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)},$$

where  $K(u) = \prod_{j=1}^k k(u_k)$ ,  $u = (u_1, \dots, u_k)'$ ,  $k(\cdot)$  is a kernel function and  $h$  denotes the bandwidth. In this application we employed the Gaussian kernel. We have plugged-in the estimate of the optimal bandwidth minimizing the integrated mean square error (see Härdle, 1990). The confidence bands for the nonparametric fits are also reported.

As an exploratory analysis, the nonparametric estimates illustrate the differences in income-elasticities across subsamples. As a decision rule, this analy-

- Nonparametric estimates and 95 % confidence level bands
- - - - Poisson
- ..... Negative Binomial
- - - - Hurdle Negative Binomial



**Fig. 2.** Estimated conditional probabilities for working men (Other variables fixed at their mean value.)

sis suggests the rejection of the models that predict outside the nonparametric confidence bands. In this sense, graphics show that estimates based on the negative binomial and hurdle models are close to the nonparametric estimates and they are generally inside the confidence intervals. The Poisson estimates are quite different and they are usually outside the confidence bands. As it is well known, this specification underestimates the conditional probabilities of zero and the highest counts, whilst, in our case, it overestimates the conditional probability of one visit. However, this graphical method cannot be used as a serious model selection criterion. Nonparametric estimates suffer from the “curse of dimensionality” and are very sensitive to selected bandwidths. Furthermore, estimates are very unstable near the data bounds, which restrict their performance as a selection rule.

Another popular model-selection method consists of comparing different marginal probability estimates computed under different model specifications with the sample frequencies of counts (see e.g. Gurmu and Trivedi, 1996 and Deb and Trivedi, 1997). That is, if  $\hat{P}_{ny}(x)$  is an estimator of  $P_y(x) = \Pr(Y = y | X = x)$ , based on some of the three alternative specifications, we compare

**Table 6.** Estimated marginal probabilities

	SAMPLE	POISSON	NEG.BIN	HURDLE
WORKING WOMEN				
$P(Y = 0)$	0.805	0.661	0.808	0.815
$P(Y = 1)$	0.099	0.265	0.101	0.097
$P(Y = 2)$	0.042	0.031	0.041	0.042
$P(Y \geq 3)$	0.054	0.043	0.050	0.046
WORKING MEN				
$P(Y = 0)$	0.859	0.859	0.856	0.862
$P(Y = 1)$	0.073	0.129	0.079	0.056
$P(Y = 2)$	0.032	0.005	0.030	0.031
$P(Y \geq 3)$	0.036	0.006	0.034	0.051

$$\hat{\mathbf{P}}_n(y) = \frac{1}{n} \sum_{i=1}^n \hat{P}_{ny}(X_i)$$

with

$$\bar{\mathbf{P}}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Y_i = y),$$

the sample frequencies of counts. The chosen model is that with  $\hat{\mathbf{P}}_n(y)$  closer to  $\bar{\mathbf{P}}_n(y)$  for all  $y$ . The estimated marginal probabilities are reported in Table 6. The negative binomial and hurdle models perform better than the Poisson under this criterion. However, we are not able to discriminate between the negative binomial and the hurdle models.

Given the correct specification of the conditional mean, overdispersion tests may be implemented to test the Poisson specification against the negative binomial. These are based on the  $t$  test of  $\delta_0 = 0$  in the least-squares regression:

$$\frac{(Y_i - \hat{\lambda}_i)^2 - Y_i}{\hat{\lambda}_i} = \delta_0 \hat{\lambda}_i + \text{error},$$

where  $\hat{\lambda}_i = \exp(\hat{\alpha} + x_i' \hat{\beta})$ ,  $\hat{\alpha}$  and  $\hat{\beta}$  are the Poisson estimates of parameters. The  $p$ -values are presented in row 3 of Table 7. In the two samples, rejection of the parameter restriction specified under the null is indicated, as expected in view of previous selection criteria.

Rejection of the Poisson model against the negative binomial specification does not imply that the negative binomial specification is the correct one. In fact, a hurdle specification, and many others, also model overdispersion. Also, note that we have performed two different tests on two specification restrictions. On one hand, we have not rejected the null hypothesis that the conditional mean is linear exponential, and, on the other hand, we have rejected the null hypothesis of no overdispersion. However, we have not jointly tested both restrictions. In row 4 of Table 7, we report the bootstrap  $p$ -values of

**Table 7.** *P*-values of tests

	WORKING WOMEN	WORKING MEN
<b>REGRESSION AND BINARY DECISION</b>		
Stute (1997) test for conditional means		
$H_0 : E(Y X) = \exp(X'\beta)$	0.182	0.980
$H_0 : E[1(Y = 0)   X] = [1 + \exp(X'\beta)]^{-1}$	0.692	0.074
*The alternatives are the negation of the null		
<b>CONDITIONAL DISTRIBUTION</b>		
Regression based score test for overdispersion		
$H_0 : \delta_0 = 0$ and $H_1 : \delta_0 > 0$	0.000	0.000
Andrews (1997) test for the conditional distribution		
$H_0 : P_y(x) = \text{Poisson}$	0.000	0.000
$H_0 : P_y(x) = \text{Negative binomial}$	0.705	0.145
$H_0 : P_y(x) = \text{Hurdle negative binomial}$	0.873	0.176
*The alternatives are the negation of the null		

the Andrews test for the Poisson specification in (1), which are very close to zero.

Finally, we implemented the Andrews test for testing the negative binomial and hurdle negative binomial specifications. The null hypotheses in this case are the conditional distribution functions presented in equations (2) and (5), respectively. To implement the test, we generated 500 parametric bootstrap samples, following the procedure described in section 3.3. The bootstrap *p*-values of the test are presented in last rows of Table 7. None of the two specifications can be rejected. Upon examination of previous model selection criteria, such results are not surprising. The bootstrap *p*-values corresponding to the negative binomial specification for the female and male samples are 0.705 and 0.145, respectively. The bootstrap *p*-values corresponding to the hurdle specification for the two samples are 0.873 and 0.176. Therefore, we conclude that, in the light of available empirical evidence, the two models are indistinguishable in this application.

## 5.2. Analysis of economic results

In this section, we summarize the main results of the estimation of dental care demand models presented in previous sections. Although previous specification procedures strongly reject the Poisson model and so the assumption of conditional equidispersion of data, they are not conclusive regarding the type of decision process governing individuals' dental care demand. Predictive capacities of the negative binomial model and the hurdle negative binomial model are very similar. Therefore, in what follows, we would analyze the economic implications of estimates by combining both specifications.

Generally speaking, coefficient estimates reveal some differences across subsamples that confirm the existence of an interaction between the explanatory variables and gender when explaining dental care use. Consistent with previous

studies, dental health indicators show the main explanatory capacity as determinants of dental care utilization. The usage of dental care is estimated to be about three times greater for people who conserve more than half (but not all) of their natural teeth than for those who conserve all their denture. However, the positive effect of *DENTH3* in the female sample is only significant at the second stage level of the hurdle model. It is worth noticing that the relationship between health and dental care is not unambiguous. Dental conditions may not only be a proxy for bad dental health, but also the consequence of dental treatment. Nonetheless, so far as dental extractions are the result of previous bad dental health, reverse causality seems to be a minor problem in this case.

Contrary to other health care utilizations, the aging process seems to be negatively related to dental care use in the female case but is not significant for men. For working women, we estimate that an increase of one year leads to a  $-0.28$  proportionate change in the expected number of dentist visits. As long as contact decision reflects patient favorable attitudes towards dental care and preventive behavior, the negative effect of age are consistent with a decline in those perceptions with age, as has been claimed in other studies (Kenkel, 1994, Mueller and Monheit, 1988).

According to our analysis, working women use, on average, more dental care if they are married than if they declare other marital status. However, this effect is not statistically significant. Concerning the standard negative binomial estimates for men, marriage has a negative impact on dentist visits. But if we look at the hurdle estimates, this negative but insignificant effect acts at the contact decision stage, whilst a positive and significant effect, at a 10 percent level, is estimated at the second stage of the decision.

The estimations provide evidence in favor of a positive relationship between years of schooling and dental care utilization. Significant estimates are only available for the working men subsample. Education may eventually be correlated with high health consciousness, which in turn stimulates preventive behavior. In the case of dental health, prevention materializes in periodical dental revisions that are likely to reduce the consequences of dental illnesses and the probability of long treatments. This fact could explain both the positive impact of schooling on the contact decision and its insignificance at the second stage of decision.

Regarding income, the significance levels and the magnitude of coefficient estimates vary according to the gender of users. Since household income enters logarithmically in the linear exponential regression model, negative binomial coefficient estimates can be interpreted as a total utilization income-elasticity. It is important to remark that income coefficient could also be reflecting an indirect effect through the likelihood of having private insurance coverage. Income-elasticity corresponding to men is 0.26, though it is not statistically different from zero for working women. Hurdle estimates point to different influences of income on the two decision-making processes. On the one hand, income performs as a positive determinant of women's access to dental care, however it does not condition men's decision at this stage. On the other hand, this variable is significant at the second stage of decision for men but not for women. In other words, once contact with dental care is established, only working women follow dental treatments independently of their family income level. Figures 1 and 2 illustrate those differences as well. In these figures, we observe that hurdle specifications offer steeper relationships between the prob-

abilities of counts and income logarithm than the negative binomial specification, when other regressors are fixed at their mean values.

Once other sources of heterogeneity are controlled for, dental care utilization is unaffected by the index of medical prices, though coefficient estimates have the expected negative sign. The only significant effect is estimated at the frequency decision of working women. The insignificance of this variable may be reflecting the limitations of an aggregate price index in this context.

Finally, dentists per population does not appear to have a significant effect in any case. This finding does not support the supply inducement hypothesis. Nonetheless, an alternative explanation for this result lies with the possibility that inducement arises through the kind of treatment received by patients. In such a case, supply inducement would affect individuals private dental expenditure but not their number of visits to the dentist. Therefore, further empirical analysis is needed in order to understand the influence of health professionals on patients' behavior.

## 6. Conclusions

In this paper we have compared the performance of several goodness-of-fit techniques for the analysis of count data. In particular, we have focused on the implementation of specification tests for the functional forms of the conditional mean and the distribution function that are consistent against nonparametric alternatives. Since the tests statistics are not distribution free, the critical values have to be approximated by bootstrap techniques. The empirical implementation was illustrated by estimating count data models of demand for dental care. The alternative specifications considered were the Poisson, negative binomial and hurdle negative binomial. The results show that the proposed specification tests are feasible tools for applied count data analysis. In the two samples considered, they recommended no rejection of the linear exponential conditional mean and no rejection of the negative binomial model. As for the economic results, the relevance of dental health, age, education and family income, as determinants of dental care utilization, is remarkable. Significant interactions of these variables with gender were also found.

## References

- [1] Álvarez B (2001) La demanda atendida de consultas médicas y servicios urgentes en España. *Investigaciones Económicas*, forthcoming
- [2] Andrews DWK (1997) A conditional Kolmogorov test. *Econometrica* 65, 5:1097–1128
- [3] Bierens HJ (1982) Consistent model specification tests. *Journal of Econometrics* 20, 1:105–134
- [4] Cameron AC, Johansson P (1997) Count data regressions using series expansions with applications. *Journal of Applied Econometrics* 12, 3:203–223
- [5] Cameron AC, Trivedi PK (1990) Regression-based tests for overdispersion in the poisson model. *Journal of Econometrics* 46, 3:347–364
- [6] Cameron AC, Trivedi PK (1998) *Regression analysis of count data*. Cambridge Econometric Society Monographs, no. 30. Cambridge, New York and Melbourne. Cambridge University Press
- [7] Cameron AC, Trivedi PK, Milne F, Piggot J (1988) A microeconomic model of the demand for health care and health insurance in Australia. *Review of Economic Studies* 55, 1:85–106

- [8] Cameron AC, Windmeijer FA (1996) *R*-squared measures for count data regression models with applications to health care utilization. *Journal of Business & Economic Statistics* 14, 2:209–220
- [9] Cox DR (1962) Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B* 17:406–424
- [10] Deb P, Trivedi PK (1997) Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics* 12, 3:313–336
- [11] Delgado MA, Kniesner TJ (1997) Count data models with variance of unknown form: An application to a hedonic model of worker absenteeism. *Review of Economics and Statistics* 79, 1:41–49
- [12] Eubank R, Spiegelman S (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association* 85, 410:387–392
- [13] Fuchs V (1982) Time preference and health: An explanatory study. In *Economic Aspects of Health*. Chicago. University of Chicago Press
- [14] Geil P, Million A, Rotte R, Zimmermann KF (1997) Economic incentives and hospitalization in Germany. *Journal of Applied Econometrics* 12, 3:295–311
- [15] Gerfin M (1996) Parametric and semi-parametric estimation of the binary response model of labor market participation. *Journal of Applied Econometrics* 11, 3:321–339
- [16] Gourieroux C, Monfort A, Trognon A (1984) Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica* 52, 3:701–720
- [17] Grossman M (1972) The demand for health: A theoretical and empirical investigation. New York. National Bureau of Economic Research
- [18] Grytten J, Holst D, Laake P (1990) Supplier inducement: Its effects on dental services in Norway. *Journal of Health Economics* 9, 4:483–491
- [19] Gurmu S (1997) Semi-parametric estimation of hurdle regression models with an application to medicaid utilization. *Journal of Applied Econometrics* 12, 3:225–242
- [20] Gurmu S, Trivedi PK (1996) Excess zeros in count models for recreational trips. *Journal of Business and Economics Statistics* 14, 4:469–477
- [21] Härdle W (1990) Applied nonparametric regression. *Econometric Society Monographs* 19. Cambridge, New York and Melbourne. Cambridge University Press
- [22] Härdle W, Mammen E (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21:1926–1947
- [23] Hong Y, White H (1995) Consistent specification testing via nonparametric series regressions. *Econometrica* 63, 5:1133–1160
- [24] Hunt-McCool J, Kiker BF, Ng YC (1994) Estimates of the demand for medical care under different functional forms. *Journal of Applied Econometrics* 9, 2:201–218
- [25] Kenkel DS (1994) The demand for preventive medical care. *Applied Economics* 26, 4:313–325
- [26] Liu R (1988) Bootstrap procedures under some non-iid models. *Annals of Statistics* 16:1697–1708
- [27] Manning WG, Phelps CE (1979) The demand for dental care. *The Bell Journal of Economics* 10, 2:503–525
- [28] Mossialos E, Le Grand J (1999) Cost containment in EU: An overview. In *Health Care and Cost Containment in the European Union*, Mossialos E, Le Grand J (eds.). Aldershot, U.K., Brookfield, Vt. and Sidney. Ashgate
- [29] Mueller CD, Monheit AC (1988) Insurance coverage and the demand for dental care. *Journal of Health Economics* 7, 1:59–72
- [30] Mullahy J (1986) Specification and testing of some modified count data models. *Journal of Econometrics* 33, 3:341–365
- [31] Mullahy J (1997) Heterogeneity, excess zeros, and the structure of count data models. *Journal of Applied Econometrics* 12, 3:337–350
- [32] Murphy KM, Topel RH (1985) Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics* 3, 4:370–379
- [33] Pohlmeier W, Ulrich V (1995) An econometric model of the two part decision process of the demand for health care. *The Journal of Human Resources* 30, 2:339–361
- [34] Sindelar JL (1982) Differential use of medical care by sex. *Journal of Political Economy* 90, 5:1003–1019
- [35] Stute W (1997) Nonparametric model checks for regression. *Annals of Statistics* 25:613–641

- [36] Stute W, González Manteiga W, Presedo Quindimil M (1998) Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* 93, 441:141–149
- [37] Wand MP, Jones MC (1995) *Kernel smoothing*. Chapman & Hall
- [38] Winkelmann R, Zimmermann KF (1995) Recent developments in count data modelling: Theory and application. *Journal of Economic Surveys* 9, 1:1–24
- [39] Wu CFJ (1986) Jack-knife, bootstrap and other resampling methods in regression analysis. (with discussion), *Annals of Statistics* 14:1261–1350
- [40] Zheng X (1998) A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory* 14, 1:123–138