

WILEY

Demand for Medical Care by the Elderly: A Finite Mixture Approach

Author(s): Partha Deb and Pravin K. Trivedi

Source: *Journal of Applied Econometrics*, May - Jun., 1997, Vol. 12, No. 3, Special Issue: Econometric Models of Event Counts (May - Jun., 1997), pp. 313-336

Published by: Wiley

Stable URL: <https://www.jstor.org/stable/2285252>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Wiley is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Applied Econometrics*

JSTOR

DEMAND FOR MEDICAL CARE BY THE ELDERLY: A FINITE MIXTURE APPROACH

PARTHA DEB^{a*} AND PRAVIN K. TRIVEDI^b

^a*Department of Economics, Indiana University-Purdue University Indianapolis, Cavanaugh Hall 516,
Indianapolis, IN 46202, USA*

^b*Department of Economics, Indiana University, Wylie Hall, Bloomington, IN 47405, USA*

SUMMARY

In this article we develop a finite mixture negative binomial count model that accommodates unobserved heterogeneity in an intuitive and analytically tractable manner. This model, the standard negative binomial model, and its hurdle extension are estimated for six measures of medical care demand by the elderly using a sample from the 1987 National Medical Expenditure Survey. The finite mixture model is preferred overall by statistical model selection criteria. Two points of support adequately describe the distribution of the unobserved heterogeneity, suggesting two latent populations, the 'healthy' and the 'ill' whose fitted distributions differ substantially from each other. © 1997 by John Wiley & Sons, Ltd.

J. Appl. Econ., 12, 313–336 (1997)

No. of Figures: 2. No. of Tables: 7. No. of References: 30.

1. INTRODUCTION

Cost and access continue to be the fundamental issues in the debate over the future of the American health-care system. An important element of this debate concerns the health-care needs of the elderly. During the past two decades, the population aged 65 and over has increased more than twice as fast as the younger population and they account for a disproportionate share of medical care expenditures.¹ This disproportionate share of expenditures can be attributed, in part, to the relatively greater need of the elderly for individualized care, and the increasing incidence of chronic diseases and impairment with age. Therefore, the increasing size and life expectancy of the elderly population have significant implications for health care: this motivates the present study.

Theoretical analyses of medical care utilization suggest two approaches. The first takes the traditional consumer theory approach (Grossman, 1972; Muurinen, 1982) and views medical care demand as primarily patient determined, though conditioned by the health-care delivery system. One-step econometric models in this tradition are estimated by Duan *et al.* (1983) and Cameron *et al.* (1988), among others. The second approach uses a principal–agent set-up (Zweifel, 1981) in which the physician (agent) determines utilization on behalf of the patient (principal) once initial contact is made. Econometric models which have an interpretation in the principal–agent tradition are estimated by Manning *et al.* (1987), Pohlmeier and Ulrich (1995), and others.

* Correspondence to: Partha Deb, Department of Economics, Indiana University-Purdue University Indianapolis (IUPUI), Cavanaugh Hall 516, 425 University Blvd, Indianapolis, IN 46205, USA. E-mail: pdeb@indyunix.iupui.edu

¹ In 1987, these elderly comprised 12% of the US population and accounted for 36% of health-care expenditures (Waldo *et al.*, 1989).

Previous analyses of the elderly by Huang and Koropecy (1976), who estimate a one-step model of utilization, Rice (1984), Christensen, Long, and Rodgers (1987), and Cartwright, Hu, and Huang (1992), who use two-step models, find that individuals with greater health-care insurance coverage tend to use more medical care than individuals with less generous coverage. Moreover, health status is often shown to be the most important determinant of utilization.

This article models counts of medical care utilization by the elderly in the United States using data from the National Medical Expenditure Survey, 1987. A feature of these data is that they include a high proportion of zero counts, corresponding to zero recorded demand over the sample interval, and a high degree of unconditional overdispersion. With these features in mind, we consider three econometric specifications: (1) the standard negative binomial specification, (2) a hurdles variant of the negative binomial—a ‘two-part model’, and (3) a finite mixture of negative binomials which provides additional flexibility compared with the hurdles model.

The two-part hurdles model has performed satisfactorily in several empirical studies (Pohlmeier and Ulrich, 1995; Gurmu, 1996) and is typically superior to standard one-part specifications. It may be interpreted as a principal–agent type model in which the first part specifies the decision to seek care as a binary outcome process, and the second part models the number of visits for the individuals who receive some care. The second part allows for population heterogeneity among the users of health care.

The finite mixture model allows for additional population heterogeneity but avoids the sharp dichotomy between the population of ‘users’ and ‘non-users’. In the formulation of this paper the underlying unobserved heterogeneity which splits the population into latent classes is assumed to be based on the person’s latent long-term health status. Proxy variables such as self-perceived health status and chronic health conditions are not necessarily good indicators of an individual’s health, and may not fully capture population heterogeneity from this source. Hence we assume that the most important source of unobserved population heterogeneity is unobserved or latent health status. In the case of a two-point finite mixture model, a dichotomy between the ‘healthy’ and the ‘ill’ groups, whose demands for health care are characterized by, respectively, low mean and low variance, and high mean and high variance may be suggested. While the two-step model captures an important feature of the data that the one-step model does not, the proposed finite mixture model has greater flexibility of functional form, but does not rely for its justification on a principal–agent type formulation.

To the best of our knowledge, no application of finite mixture models in health economics exists. Yet the intuitive appeal of a finite mixture specification as a plausible alternative to the existing models cannot be ignored. Finite mixture and latent class analysis, a closely related area, have received increasing attention in the statistics literature mainly because of the number of areas in which such distributions are encountered (see McLachlan and Basford, 1988, and Lindsay, 1995, for numerous applications). Econometric applications of finite mixture models include the seminal work of Heckman and Singer (1984), of Wedel *et al.* (1993) to marketing data and El-Gamal and Grether (1995) to data from experiments in decision making under uncertainty.

In the following section of the paper, we present several mixture count models used in this paper. Section 3 discusses the model comparison, selection and evaluation strategies used in this article. Section 4 reports a small simulation investigation of several of these criteria. The data are described in Section 5. The empirical results of the specification tests and the evaluation and interpretation of the empirical results are in Section 6. Section 7 concludes the article.

2. MIXTURE MODELS FOR COUNTS

Consider a single probabilistic health event, e.g. sickness or health maintenance, which may or may not result in the demand for some type of health care, e.g. doctor consultation. The decision to seek health care may be treated within the random utility framework used in binary choice models. Denote by U_0 the utility of not seeking care, by U_1 the utility of seeking care. Let $U_{1i} = x_i'\beta_1 + \varepsilon_{1i}$, let $U_{0i} = x_i'\beta_0 + \varepsilon_{0i}$ where x_i is the vector of individual attributes, and ε_{1i} and ε_{0i} are random errors. Then for individual i who seeks care, we have

$$\begin{aligned} U_{1i} &> U_{0i} \\ p_{10} &< x_i'(\beta_1 - \beta_0) \end{aligned}$$

where $p_{10} = \Pr[\varepsilon_{0i} - \varepsilon_{1i}]$. Thus, the decision to seek care is characterized by the binomial model as is standard in a binary outcome model.

For the population, consider N independent Bernoulli trials occurring over some time interval. By the 'Law of Rare Events', the distribution of the number of successes can be shown to follow the Poisson law. Formally, let $Y_{n,p_{10}}$ denote the total number of successes in a large number N of independent Bernoulli trials with success probability p_{10} of each trial being small. Then

$$\Pr[Y_{N,p_{10}} = k] = \binom{N}{k} (p_{10})^k (1 - p_{10})^{N-k} \quad k = 0, 1, \dots, N$$

In the limiting case where $N \rightarrow \infty$, $p_{10} \rightarrow 0$, and $Np_{10} = \mu > 0$, i.e. the average μ is held constant while $N \rightarrow \infty$, then

$$\lim_{N \rightarrow \infty} \left[\binom{N}{k} \left(\frac{\mu}{N}\right)^k \left(1 - \frac{\mu}{N}\right)^{N-k} \right] = \frac{\mu^k e^{-\mu}}{k!}$$

for $k \geq 0$, which is the Poisson probability distribution with parameter μ .

The above argument justifies the framework of count data models for the study of medical care utilization based on event counts. It can be generalized in a straightforward manner to allow for serial correlation or unobserved heterogeneity which will imply an overdispersed count model like the negative binomial. Since empirical evidence (Cameron and Trivedi, 1986; Pohlmeier and Ulrich, 1995) suggests that overdispersion is a common feature of medical care utilization data in this article we take the negative binomial family as the benchmark and consider generalizations based on it.

2.1. Negative Binomial (NB)

Let y_i be a count dependent variable that takes values 0, 1, 2, The density function for the negative binomial (NB) model is given by

$$f(y_i) = \frac{\Gamma(y_i + \psi_i)}{\Gamma(\psi_i)\Gamma(y_i + 1)} \left(\frac{\psi_i}{\lambda_i + \psi_i}\right)^{\psi_i} \left(\frac{\lambda_i}{\lambda_i + \psi_i}\right)^{y_i} \quad (1)$$

where

$$\lambda_i = \exp(x_i'\beta)$$

and the precision parameter (ψ_i^{-1}) specified as

$$\psi_i = (1/\alpha)\lambda_i^k$$

where $\alpha > 0$ is an overdispersion parameter and k is an arbitrary constant. In this specification, the mean function is

$$E(y_i | x_i) = \lambda_i \quad (2)$$

and the variance function is

$$V(y_i | x_i) = \lambda_i + \alpha\lambda_i^{2-k} \quad (3)$$

If $\alpha = 0$ we get the Poisson model. The negative binomial-1 (NB1) model is obtained by specifying $k = 1$ while the negative binomial-2 (NB2) is obtained by setting $k = 0$. Models with k estimated along with other parameters are not considered here.

2.2. Zero-inflated NB (ZINB)

The excess zeros feature of the data can be accommodated by a zero-inflated variant of the standard Poisson and negative binomial models. For example, the zeros may come partly from some NB model, say with probability π , and partly, with probability $1 - \pi$, from the population of 'non-users'. Thus the model permits mixing with respect to the zero counts in a different way from that for positive counts. A ZINB specification such as this is usually justified in terms of a sampling frame that samples event counts both from the population who are 'at risk' and those 'not at risk'. The ZINB model is not considered here because it is a special (degenerate) case of the finite mixture negative binomial model that permits mixing with respect to both zero and positive counts, which will be considered.

2.3. Hurdle Negative Binomial (HNB)

In the hurdles approach to the 'excess' zeros problem, the process that determines the zero/non-zero count threshold is different from the process that determines the count once the hurdle is crossed. In the case of markets for health care, the motivation for hurdle models comes from principal-agent theories of demand which suggest that the patient initially chooses whether to seek treatment, but once in treatment the physician determines the number of visits. The first part of the hurdle (HNB) two-part structure is specified as

$$\Pr_h(y_i = 0 | x_i) = \left(\frac{\psi_{h,i}}{\lambda_{h,i} + \psi_{h,i}} \right)^{\psi_{h,i}} \quad (4)$$

$$\Pr_h(y_i > 0 | x_i) = 1 - \left(\frac{\psi_{h,i}}{\lambda_{h,i} + \psi_{h,i}} \right)^{\psi_{h,i}} \quad (5)$$

where the subscript h denotes parameters associated with the 'hurdle distribution'. The likelihood function associated with this stage of the hurdle process can be maximized independently of the specification of the second stage. This stage of the hurdle process involves only binary

information so the parameters (β) of the mean function and the parameter α are not separately identifiable. In this application we fix the value of α to 1: this is a sufficient condition for identification as long as the model is otherwise identified.

Once the hurdle is crossed, the data are assumed to follow the density for a truncated negative binomial distribution given by

$$f(y_i | x_i, y_i > 0) = \frac{\Gamma(y_i + \psi_i)}{\Gamma(\psi_i)\Gamma(y_i + 1)} \left[\left(\frac{\lambda_i + \psi_i}{\psi_i} \right)^{\psi_i} - 1 \right]^{-1} \left(\frac{\lambda_i}{\lambda_i + \psi_i} \right)^{y_i} \quad (6)$$

This stage is estimated using the subsample of observations with positive values of y_i . Denote

$$\Pr(y_i > 0 | x_i) = 1 - \left(\frac{\psi_i}{\lambda_i + \psi_i} \right)^{\psi_i} \quad (7)$$

then, in the HNB model, the mean of the count variable is given by

$$E(y_i | x_i) = \frac{\Pr_h(y_i > 0 | x_i)}{\Pr(y_i > 0 | x_i)} \lambda_i \quad (8)$$

and the variance by

$$V(y_i | x_i) = \frac{\Pr_h(y_i > 0 | x_i)}{\Pr(y_i > 0 | x_i)} \left[\lambda_i + \alpha \lambda_i^{2-k} + \left(1 - \frac{\Pr_h(y_i > 0 | x_i)}{\Pr(y_i > 0 | x_i)} \right) \lambda_i^2 \right] \quad (9)$$

Note that both the mean and the variance in the hurdle model are, in general, different from their standard-model counterparts.

2.4. C-point Finite Mixture Negative Binomial (FMNB-C)

In a finite mixture model a random variable is postulated as a draw from a superpopulation which is an additive mixture of C distinct populations in proportions π_1, \dots, π_C , where $\sum_{j=1}^C \pi_j = 1$, $\pi_j > 0$ ($j = 1, \dots, C$). The density is given by

$$f(y_i | \Theta) = \sum_{j=1}^{C-1} \pi_j f_j(y_i | \theta_j) + \pi_C f_C(y_i | \theta_C) \quad (10)$$

where the mixing probabilities π_j are estimated along with all other parameters, denoted Θ . Also $\pi_C = (1 - \sum_{j=1}^{C-1} \pi_j)$. For identifiability, we use the condition that $\pi_1 \geq \pi_2 \geq \dots \geq \pi_C$ which can always be satisfied by rearrangement, post estimation (McLachlan and Basford, 1988).

The component distributions in a C -point finite mixture negative binomial model (FMNB-C) are specified as

$$f_j(y_i) = \frac{\Gamma(y_i + \psi_{j,i})}{\Gamma(\psi_{j,i})\Gamma(y_i + 1)} \left(\frac{\psi_{j,i}}{\lambda_{j,i} + \psi_{j,i}} \right)^{\psi_{j,i}} \left(\frac{\lambda_{j,i}}{\lambda_{j,i} + \psi_{j,i}} \right)^{y_i} \quad (11)$$

where $j = 1, 2, \dots, C$ are the latent classes, $\lambda_{j,i} = \exp(x_i' \beta_j)$ and $\psi_{j,i} = (1/\alpha_j) \lambda_{j,i}^k$.

If $\dim[\beta_j] = k$, then the fully unconstrained C component mixture has dimension $C(k + 1) - 1$. In practice considerations of parsimonious parameterizations may motivate restrictions across the component densities. Therefore, we also consider a constrained FMNB-C (CFMNB-C) with all slope parameters in β_j restricted to be equal across all C components. The differences between component distributions then arise only from intercept differences.

The finite mixture framework offers a number of advantages. It has a natural representation of heterogeneity in a finite, usually small, number of latent classes, each of which may be regarded as a 'type'. The choice of the number of components in the mixture determines the number of 'types', but the choice of the functional form for the density can accommodate heterogeneity within each component. Thus the formulation is flexible enough to allow for a continuum of states. Second, the finite mixture approach is semi-parametric: it does not require any distributional assumptions for the mixing variable. Third, the results of Laird (1978) and Heckman and Singer (1984) suggest that estimates of such finite mixture models may provide good numerical approximations even if the underlying mixing distribution is continuous. Finally, the choice of a continuous mixing density for some parametric count models is sometimes restrictive and computationally intractable except when the conditional (kernel) and mixing densities are from conjugate families, since otherwise the marginal density will not have a closed form. By contrast, there are several promising approaches available for estimating finite mixture models (Böhning, 1995).

2.4.1. Latent class analysis

The finite mixture model is closely related to latent class analysis (Aitkin and Rubin, 1985; Wedel *et al.*, 1993). Let $\mathbf{d}_i = (d_{i1}, \dots, d_{iC})$, define a $(C \times 1)$ vector of an indicator (0/1 dummy) variables such that $d_{ij} = 1$, $\sum_{j=1}^C d_{ij} = 1$, indicates that y_i was drawn from the j th (latent) group or class for $i = 1, \dots, n$. That is, each observation may be regarded as a draw from one of the C latent classes or 'types', each with its own distribution. The latent class model specifies that $(y_i | \mathbf{d}_i)$ are independently distributed with densities

$$\sum_{j=1}^C d_{ij} f_j(y_i | \theta_j) = \prod_{j=1}^C f_j(y_i | \theta_j)^{d_{ij}}$$

and $(\mathbf{d}_i | \boldsymbol{\pi}, \boldsymbol{\Theta})$ are iid with multinomial distribution

$$\prod_{j=1}^C \pi_j^{d_{ij}} \quad 0 < \pi_j < 1 \quad \sum_{j=1}^C \pi_j = 1$$

where $\boldsymbol{\pi}' = (\pi_1, \pi_2, \dots, \pi_C)$, $\boldsymbol{\Theta} = [\theta_1, \theta_2, \dots, \theta_C]$. The last two relations lead to the likelihood function

$$L(\boldsymbol{\pi}, \boldsymbol{\Theta} | \mathbf{y}) = \prod_{i=1}^n \sum_{j=1}^C \pi_j^{d_{ij}} (f_j(\mathbf{y}; \theta_j))^{d_{ij}} \quad 0 < \pi_j < 1 \quad \sum_{j=1}^C \pi_j = 1 \quad (12)$$

An alternative formulation of this likelihood may be based on equation (10).

The results reported in this article are based on the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm which maximizes the likelihood directly based on equation (10) using only

likelihood scores. The algorithm, allowing for constraints $0 \leq \pi_j \leq 1$, programmed in SAS/IML, is a variable metric quasi-Newton algorithm. Upon convergence, robust standard errors of the parameter estimates are computed using the usual sandwich formula. An alternative widely recommended procedure for estimating the finite mixture model is the EM algorithm which is quite slow to converge (Brännäs and Rosenqvist, 1994; Böhning, 1995).

2.4.2. *Properties of finite mixtures*

An observation in each latent class is a draw from a component distribution. The component densities are assumed to differ in location via the conditional mean function and in scale via the dispersion parameter. Wedel *et al.* (1993) proposed a finite mixture Poisson model for estimating the effects of direct marketing on purchases. Their approach also allows for heterogeneity in the conditional mean functions of the component densities. This article uses the more general negative binomial densities which additionally allow for differences in overdispersion within classes. We also give greater consideration to diagnostics and testing.

Since in a FMNB-C model mixing is allowed with respect to *all* observations, and not only with respect to zero observations as in ZINB, the former may also be considered a generalization of the latter. In ZINB there is a sharp and unrealistic distinction between the healthy (not ill) population with zero demand, and the ill population with positive demand which is avoided by the FMNB-C model.

The factorial raw moments of a finite mixture distribution may be derived using the general formula

$$m'_r(y) = E[y^r] = \sum_{j=1}^C \pi_j m'_r[y | f_j(y)]$$

where m'_r denotes the r th factorial raw moment and $f_j(y)$ the j th component (Johnson, Kotz, and Kemp, 1992), whence the central moments can then be derived.

The mean and variance are given, respectively, by

$$E(y_i | x_i) = \bar{\lambda}_i = \sum_{j=1}^C \pi_j \lambda_i \quad (13)$$

and

$$V(y_i | x_i) = \sum_{j=1}^C (\pi_j \lambda_{j,i} [1 + \alpha_j \lambda_{j,i}^{-k}]) + \bar{\lambda}_i - \bar{\lambda}_i^2 \quad (14)$$

It is interesting to note that the CFMNB-C may be consistently estimated, upto the intercept, by Poisson maximum likelihood. This follows from the previous expression for $E(y_i | x_i)$ which, given the equality of the slope parameters, is also the conditional mean for the Poisson model. However, the intercept of the constrained finite mixture is a weighted sum of the intercepts in the components, with weights being the π_i . Hence Poisson maximum likelihood will yield only an estimate of the weighted sum of the intercepts and cannot identify the individual components which are often of economic interest.

If π_j are given, the posterior probability that observation y_i belongs to the population $j, j = 1, 2, \dots, C$, is given by

$$\Pr[y_i \in \text{population } j] = \frac{\pi_j f_j(y_i | x'_i, \theta_j)}{\sum_{j=1}^C \pi_j f_j(y_i | x'_i, \theta_j)} \quad (15)$$

While the HNB and FMNB-C models both allow for greater flexibility relative to the NB model, the process that generates the data in each case is different. In the hurdle model, the underlying structure splits the population into two groups of 'users' (or 'at risk' population) and 'non-users' ('not at risk' population). Then *only* the users of care determine the number of visits. Post estimation of HNB, the probability that an individual belongs to the 'non-user' group may be obtained by subtracting the estimated proportion of zero counts due to the 'users' from the observed proportion of zeros.

The sharp dichotomy between 'users' and 'non-users' may not be tenable in the case of demand for health care, where even a healthy individual will not typically be a 'non-user' since many forms of health care among the elderly will take the form of health maintenance and precautionary care (Grossman, 1972). A more tenable distinction is between light and heavy users of health care, determined by health status, attitudes to health risk, and a choice of lifestyle. The finite mixture model, in which there is *no* distinction between users and non-users of care, but which can distinguish between groups with high average demand and low average demand therefore provides a better framework.

3. SPECIFICATION TESTING AND MODEL SELECTION

Two important issues in evaluating fitted finite mixture models are first, whether there is evidence for the presence of more than one components, i.e. mixing is present, and second, whether a global maximum is attained in estimation. Lindsay and Roeder (1992) have developed diagnostic tools for checking these properties for the case of exponential mixtures. Here we apply these tools in a heuristic fashion to the FMNB-C models which are not one-parameter exponential mixtures.

3.1. Tests of Presence of Mixture

Lindsay and Roeder present two diagnostic tools, the residual function and the gradient function. Of these the second, the directional gradient function,

$$d(y, \hat{C}, \theta^*) = \left(\frac{\hat{p}(y; \theta^*)}{\hat{p}(y; \hat{C})} - 1 \right) \quad y = 0, 1, 2, \dots \quad (16)$$

where θ^* is the unicomponent MLE and \hat{C} is the C component MLE, is more useful because it is smoother than the residual function. Following Lindsay and Roeder (1992) the graph of equation (16) is used as a diagnostic 'test' for the presence of a mixture. The convexity of the graph is interpreted as evidence in favour of a mixture. However, note that such convexity may be observed for more than one value of C , leaving open the issue of which value to select for C .

The use of the weighted sum measure

$$D(\hat{C}, \theta^*) = \sum_{y \in S} d(y, \hat{C}, \theta^*) p(y) \quad (17)$$

as an additional diagnostic is also suggested but we do not report this as it was found difficult to interpret. The poor discriminatory content of this statistic is not unexpected. Lindsay and Roeder argue that a truncated gradient statistic is preferred for unbounded densities, precisely because the unadjusted gradient statistic is likely to be uninformative. However they provide little guidance on how the statistic should be truncated so we do not pursue that issue in this article.

3.2. Model-selection Tests

Given the presence of mixture, one may either sequentially compare models with different values of C , or compare unconstrained and constrained models for a given C . Though these are both nested hypotheses, the use of the likelihood ratio (LR) test is only appropriate for the latter simplification, not the former in which the hypothesis is on the boundary of the parameter space and which violates the standard regularity conditions for maximum likelihood. For example, in a model with no component-density parameters estimated, the LR test of the null hypothesis $H_0 : \pi_C = 0$ versus $H_a : \pi_C \neq 0$, does not have the usual null $\chi^2(1)$ distribution. Instead, the asymptotic distribution of the likelihood ratio is a weighted χ^2 . Böhning *et al.* (1994) show, using simulation analysis of the distribution of the likelihood ratio test in for several distributions involving a boundary hypothesis, that the use of the nominal $\chi^2(1)$ test is likely to underreject the false null. Therefore, systematic reliance on the LR test may cause an investigator to choose a value of C that is too small. However, the use of information criteria (IC)—we use the Akaike information criterion (AIC) and the Bayesian information criterion (BIC)—for model selection has formal justification (Leroux, 1992).²

Model simplification in going from HNB to NB involves a standard nested hypothesis, hence it may be based on the LR test. Finally, if FMNB- C , $C > 1$, and HNB are the preferred models after initial tests, one will wish to select between these two. Again this involves non-nested comparisons, for which we use IC to choose between them.

Schematically, the strategy described above is given by:

- FMNB- $C \xrightarrow{IC} \text{FMNB-1} \equiv \text{NB}$
- FMNB- $C \xrightarrow{LR} \text{CFMNB-}C$
- HNB $\xrightarrow{LR} \text{NB}$
- FMNB- $C \xrightarrow{IC} \text{HNB}$.

Note that NB in the above may refer to either NB1 or NB2.

Finally, to evaluate the goodness of fit of the model selected after this multi-level comparison of models, the chi-square diagnostic test $(\mathbf{f} - \hat{\mathbf{f}})' \hat{\Sigma}^{-1} (\mathbf{f} - \hat{\mathbf{f}})$ is used where $\mathbf{f} - \hat{\mathbf{f}}$ is the q -dimensional vector of difference between sample and fitted cell frequencies, q is the number of cells used in the test, and $\hat{\Sigma}$ is the estimated variance matrix of the difference. Under the null hypothesis of no

² Leroux proves that under regularity conditions the maximum penalized likelihood approach, e.g. the use of Akaike and Bayes information criteria, lead to a consistent estimator of the true finite mixture model (Leroux, 1992, Section 3.3).

misspecification the test has an asymptotic $\chi^2(q-1)$ distribution (Heckman, 1984; Andrews, 1988). In general the last cell aggregates over several sample support points. If the sample mean is low, then fewer support points would be used than if the value is large. The covariance matrix of $\mathbf{f} - \hat{\mathbf{f}}$ is estimated by an auxiliary regression based on the outer product of gradients suggested by Andrews (1988, Appendix 5), because of the relative computational simplicity of this method compared with the alternative more complicated expressions (Andrews, 1988).

4. A SIMULATION EXPERIMENT

This section briefly summarizes the results of a small Monte Carlo experiment in which the finite sample properties of the various tests used in this article (details are available upon request). The main motivation for this section comes from an absence of evidence on the comparative performance of the tests, several of which are known to have case-specific properties. Specific issues of interest are the properties of the tests used for model comparison and evaluation, and the ability of the test procedures to discriminate between the HNB and the FMNB-2 model when the true model may be either.

The data-generating process is either NB, HNB, or FMNB-2, within the NB2 family. Samples of 500 NB2 random variates are drawn by constructing a mixture of Poisson and Gamma random variates. The mean equation for the NB, and for the components of HNB and FMNB-2, are specified as consisting of an intercept and one uniformly distributed covariate. The parameter values generate count data consistent with two stylized facts of medical care utilization: relatively small means—the means are between 0.5 and 2, and a large proportion of zeros—there are between 40% and 70% zeros. For the finite mixture model, the mixing probability is varied between 0.7 and 0.9. For each data-generating process, parameters of NB, HNB, and FMNB-2 models are estimated by maximum likelihood so that for each *dgp*, one estimated model is the true model while the other two models are misspecified. AIC, BIC, the χ^2 goodness-of-fit test, and LR tests for NB versus HNB, and for NB versus FMNB-2, are computed in each of 500 replications for each experimental design point.

The results show that the LR test for NB versus HNB has the correct size under the null and quite good power under the alternative. Second, if one uses the usual χ^2 critical value for the LR test of NB versus FMNB-2, the test is undersized. However, the test has excellent power under the alternative even when the usual critical value is used. When the difference in component means is about 2, the test rejects over 90% of the time.

The AIC and BIC are quite successful at discriminating between the HNB and FMNB-2 models. As one might expect, the BIC is quite conservative; it tends to select the NB model fairly frequently (up to 60% of the time) when the true model is either HNB or FMNB-2. Consequently, selection of either the finite mixture or hurdle specifications by the BIC may be taken as strong evidence that the NB model is inadequate. Finally, while the goodness-of-fit test generally has the nominal size for NB, it tends to underreject (0.05 for a nominal 0.10 test) for the HNB and overreject (0.20 for a nominal 0.10 test) for FMNB-2. Overall, a failure to reject the null using the goodness-of-fit test appears to be convincing evidence that the model is correctly specified.

5. DATA AND SUMMARY STATISTICS

The data are obtained from the National Medical Expenditure Survey (NMES) which was conducted in 1987 and 1988 to provide a comprehensive picture of how Americans use and pay

for health services. The NMES is based upon a representative, national probability sample of the civilian, non-institutionalized population and individuals admitted to long-term care facilities during 1987. Under the household survey of the NMES, more than 38 000 individuals in 15 000 households across the United States were interviewed quarterly about their health insurance coverage, the services they used, and the cost and source of payments of those services. These data were verified by cross-checking information provided by survey respondents with providers of health-care services. In addition to health-care data, NMES provides information on health status, employment, sociodemographic characteristics, and economic status.

In this paper we consider a subsample of individuals ages 66 and over (a total of 4406 observations) all of whom are covered by Medicare, a public insurance programme that offers substantial protection against health care costs.³ Residents of the United States are eligible for Medicare coverage at age 65. Some individuals start receiving Medicare benefits a few months into their 65th year primarily because they fail to apply for coverage at the appropriate time. Virtually all individuals who are 66 or older are covered by Medicare. In addition, most individuals make a choice of supplemental private insurance coverage shortly before or in their 65th year because the price of such insurance rises sharply with age and coverage becomes more restrictive. Therefore, given our choice of sample, the treatment of private insurance status as predetermined, rather than endogenous, is justified.⁴ On the other hand, to the extent that private insurance purchase is in anticipation of required health care, given health status, private insurance status may be treated as endogenous. Since our specification controls for health status by including several health-status variables, the force of this argument is reduced. Exogeneity of insurance is a major econometric simplification.

We consider six mutually exclusive measures of utilization: visits to a physician in an office setting (OFP), visits to a non-physician in an office setting (OFNP), visits to a physician in a hospital outpatient setting (OPP), visits to a non-physician in an outpatient setting (OPNP), visits to an emergency room (EMR), and number of hospital stays (HOSP). Means and standard deviations of these variables are presented in Table I along with the proportions of zero values. As one might expect, most elderly persons see a physician at least once a year and a considerably smaller fraction seek care in an emergency room or are admitted to a hospital, even as outpatients.

Definitions and summary statistics for the explanatory variables are presented in Table I. The health measures include self-perceived measures of health (*EXCLHLTH* and *POORHLTH*), the number of chronic diseases and conditions (*NUMCHRON*), and a measure of disability status (*ADLDIFF*). In order to control for regional differences we use *NOREAST*, *MIDWEST* and *WEST*. The demographic variables include *AGE*, race (*BLACK*), sex (*MALE*), marital status (*MARRIED*), and education (*SCHOOL*). Finally, the economic variables are family income (*FAMINC*), employment status (*EMPLOY*), supplementary private insurance status (*PRIVINS*), and public insurance status (*MEDICAID*).

Both *PRIVINS* and *MEDICAID* serve as indicators of the price of service. However, while Medicaid coverage was quite uniform across states in 1987, i.e. most individuals covered by Medicaid had identical deductibles, coinsurance rates and covered benefits, private insurance differed widely in deductibles, coinsurance rates, and defined benefits. Therefore, while the

³ In 1990 Medicare benefit payments exceeded \$100 billion.

⁴ Cartwright, Hu, and Huang (1992) use a Wu–Hausman test for the exogeneity of supplemental private insurance in a utilization regression for a sample of individuals ages 65 and over and find that insurance status is exogenous.

Table I. Variable definitions and summary statistics

Variable	Definition	Mean	SD
<i>OFF</i>	# of physician office visits	5.77 (15.5)	6.76
<i>OFNP</i>	# of non-physician office visits	1.62 (68.2)	5.32
<i>OPP</i>	# of physician hospital outpatient visits	0.75 (77.1)	3.65
<i>OPNP</i>	# of non-physician hospital outpatient visits	0.54 (83.8)	3.88
<i>EMR</i>	# of emergency room visits	0.26 (81.8)	0.70
<i>HOSP</i>	# of hospital stays	0.30 (80.4)	0.75
<i>EXCLHLTH</i>	= 1 if self-perceived health is excellent	0.08	0.27
<i>POORHLTH</i>	= 1 if self-perceived health is poor	0.13	0.33
<i>NUMCHRON</i>	# of chronic conditions (cancer, heart attack, gall bladder problems, emphysema, arthritis, diabetes, other heart disease)	1.54	1.35
<i>ADLDIFF</i>	= 1 if the person has a condition that limits activities of daily living	0.20	0.40
<i>NOREAST</i>	= 1 if the person lives in northeastern US	0.19	0.39
<i>MIDWEST</i>	= 1 if the person lives in the midwestern US	0.26	0.44
<i>WEST</i>	= 1 if the person lives in the western US	0.18	0.39
<i>AGE</i>	age in years (divided by 10)	7.40	0.63
<i>BLACK</i>	= 1 if the person is African American	0.12	0.32
<i>MALE</i>	= 1 if the person is male	0.40	0.49
<i>MARRIED</i>	= 1 if the person is married	0.55	0.50
<i>SCHOOL</i>	# of years of education	10.3	3.74
<i>FAMINC</i>	family income in \$10 000	2.53	2.92
<i>EMPLOYED</i>	= 1 if the person is employed	0.10	0.30
<i>PRIVINS</i>	= 1 if the person is covered by private health insurance	0.78	0.42
<i>MEDICAID</i>	= 1 if the person is covered by Medicaid	0.09	0.29

Numbers in parentheses are the frequency of zero counts.

dummy variable for Medicaid is likely to measure the difference in prices for those enrolled versus others, the dummy variable for private insurance can be, at best, seen as a noisy indicator of the price of medical care for those with coverage versus those without. However, the interpretation of *MEDICAID* is complicated by the fact that it is a programme that provides health insurance for the poor. Hence, in addition to the price effect on medical care demand, *MEDICAID* is likely to capture income (poverty) effects.

6. RESULTS

6.1. Model Comparison and Evaluation

For expositional ease, Table II collects the acronyms for the various models along with brief descriptions and number of parameters estimated in each. Note that the number of parameters approximately doubles for HNB and FMNB-2 models relative to NB and CFNB-2, thus increasing the chance of overparameterization.

Table III presents likelihood ratio tests of NB versus HNB, the unicomponent model versus the two-component models (CFMNB-2 and FMNB-2) as well as the two-component models versus the corresponding three-component models. The NB model is rejected in favour of HNB in every case. The NB model is also rejected in favour of two-component mixture models, notwithstanding the fact that we have used conservative critical values. This evidence is corroborated by

Table II. Model description and number of estimated parameters

Model	Description	Parameters
NB	Negative binomial count model	18
HNB	Hurdle variant of a negative binomial model	35
CFMNB-2	Constrained two-point finite mixture of negative binomials: slope parameters equal across component densities	21
FMNB-2	Two-point finite mixture of negative binomials	37
CFMNB-3	Constrained three-point finite mixture of negative binomials: slope parameters equal across component densities	24
FMNB-3	Three-point finite mixture of negative binomials	56

Table III. Likelihood ratio tests

Null	Alternative	<i>OFP</i>	<i>OFNP</i>	<i>OPP</i>	<i>OPNP</i>	<i>EMR</i>	<i>HOSP</i>
NB1							
NB	HNB	59.8 ^a	62.4 ^a	124.0 ^a	51.4 ^a	49.9 ^a	20.8 ^a
NB	CFMNB-2	116.7 ^a	96.4 ^a	256.9 ^a	267.3 ^a	24.7 ^a	9.7 ^a
NB	FMNB-2	116.9 ^a	121.4 ^a	330.7 ^a	307.1 ^a	50.3 ^a	26.8
CFMNB-2	FMNB-2	50.2 ^a	25.0	73.8 ^a	39.7 ^a	25.6	17.1
CFMNB-2	CFMNB-3	0.002	0.01	0.001	0.78	1e-5	1e-4
FMNB-2	FMNB-3	11.3	0.01	0.001	0.01	11.8	32.6 ^a
NB2							
NB	HNB	183.4 ^a	178.0 ^a	89.6 ^a	178.2 ^a	13.4	30.8 ^a
NB	CFMNB-2	106.7 ^a	127.5 ^a	113.8 ^a	245.4 ^a	1.3	13.2 ^a
NB	FMNB-2	136.0 ^a	189.8 ^a	141.3 ^a	297.3 ^a	27.1	28.9
CFMNB-2	FMNB-2	29.2	62.3 ^a	27.5 ^a	51.7 ^a	25.7	15.7
CFMNB-2	CFMNB-3	0.003	0.001	0.003	13.6 ^a	1.2	2e-4
FMNB-2	FMNB-3	0.002	0.0002	29.9	28.7	0.004	0.002

Notes:

The test of NB versus HNB follows a $\chi^2(17)$ distribution.

The critical value for the test of NB versus CFMNB-2 is from a $\chi^2(3)$ distribution.

The critical value for the test of NB versus FMNB-2 is from a $\chi^2(19)$ distribution.

The test of CFMNB-2 versus FMNB-2 follows a $\chi^2(16)$ distribution.

The critical value for the test of CFMNB-2 versus CFMNB-3 is from a $\chi^2(3)$ distribution.

The critical value for the test of FMNB-2 versus FMNB-3 is from a $\chi^2(19)$ distribution.

^a Statistically significant at the 5% level.

the directional gradient function evaluated against FMNB-2 which are presented in Figure 1 for the NB1 specification. Similar figures for CFMNB-2 and the specifications using NB2 and were obtained but are not reported for space reasons. They appear to satisfy the convexity requirement, most clearly for OFP, and least clearly for EMR and HOSP.

Though, as already noted, there are problems with the LR test for choosing between two- and three-component mixtures, it is interesting to note that, whether based on the NB1 or NB2 mixtures, we get a significant LR statistic in only two cases out of 24. Finally, within the two-component models, the evidence in favour of the constrained model is mixed. The LR test of CFMNB-2 against FMNB-2 model, based on the NB1 specification, rejects the null model in

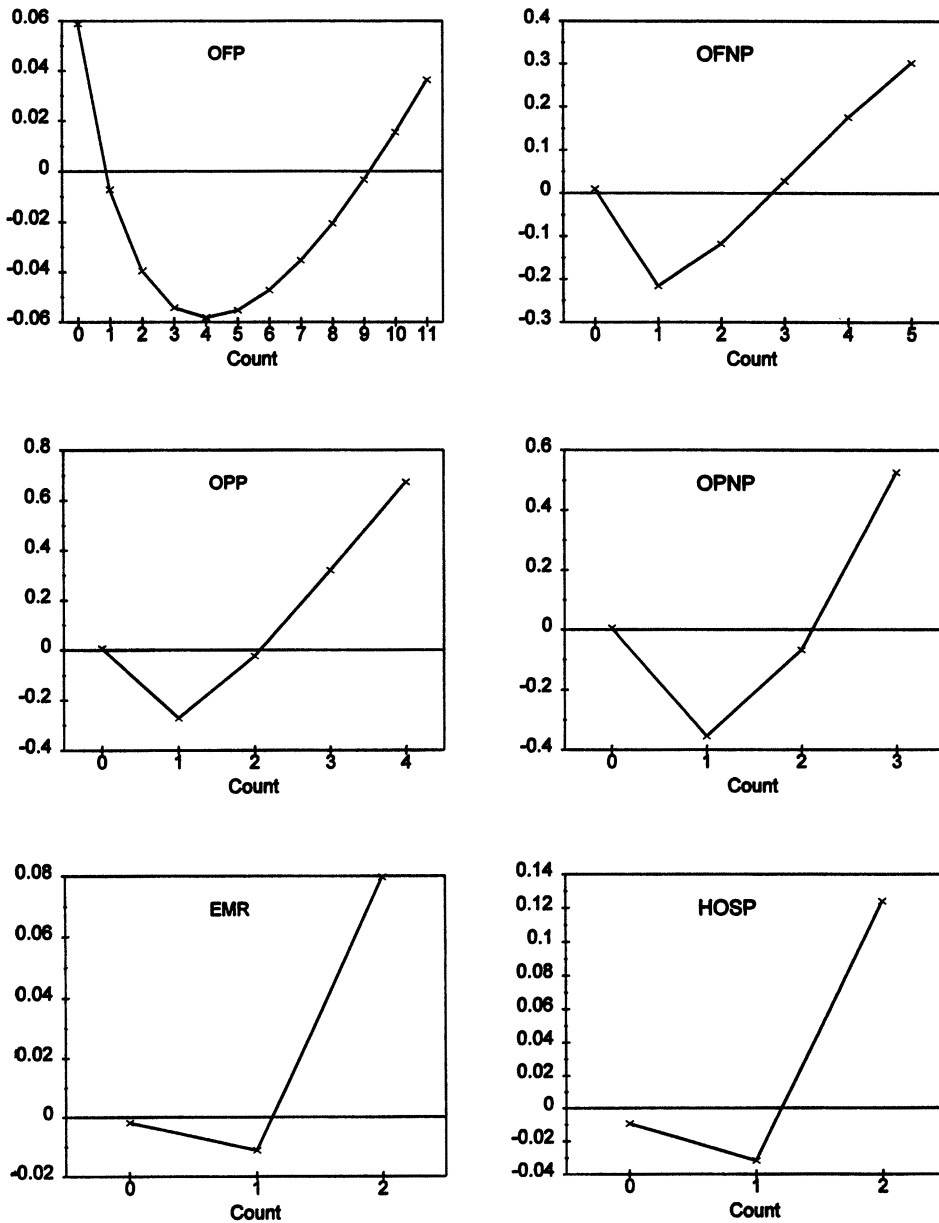


Figure 1. Directional Gradients against FMNB-2 NB1

three cases (OFF, OPP, and OPNP). If we base comparisons on the NB2 specification, the test is also rejected three times (OFNP, OPP, and OPNP).

Table IV presents values of the AIC and BIC. These show that either CFMNB-2 or FMNB-2, based on the NB1 specifications are the preferred models overall. Within the NB2 class, again

Table IV. Information criteria (AIC and BIC)

		<i>OFP</i>	<i>OFNP</i>	<i>OPP</i>	<i>OPNP</i>	<i>EMR</i>	<i>HOSP</i>
NB1							
NB	AIC	24348	11830	8409	6251	5384	5718
	BIC	24463	11945	8524	6366	5499 ^c	5833 ^{c,d}
HNB	AIC	24323	11801	8319	6234	5368	5731
	BIC	24546	12025	8543	6457	5591	5955
CFMNB-2	AIC	24238	11739 ^{a,b}	8158	5990	5365 ^{a,b}	5714 ^{a,b}
	BIC	24372 ^{c,d}	11874 ^{c,d}	8292 ^c	6124 ^{c,d}	5499	5848
FMNB-2	AIC	24220 ^{a,b}	11746	8117 ^{a,b}	5982 ^{a,b}	5371	5729
	BIC	24456	11983	8353	6218	5608	5965
CFMNB-3	AIC	24244	11745	8164	5995	5371	5720
	BIC	24397	11899	8318	6148	5524	5873
FMNB-3	AIC	24246	11784	8155	6020	5397	5734
	BIC	24604	12142	8512	6378	5755	6092
NB2							
NB	AIC	24440	11900	8229	6245	5360	5729
	BIC	24555	12015	8344	6360	5475 ^{c,d}	5844 ^c
HNB	AIC	24291 ^a	11756	8173	6101	5381	5732
	BIC	24515	11980	8397	6324	5605	5956
CFMNB-2	AIC	24340	11779	8121 ^a	6005	5365 ^a	5722 ^a
	BIC	24474 ^c	11913 ^c	8255 ^{c,d}	6140 ^c	5499	5856
FMNB-2	AIC	24342	11749 ^a	8126	5986 ^a	5371	5738
	BIC	24579	11985	8362	6222	5608	5974
CFMNB-3	AIC	24346	11785	8127	5370	5998	5728
	BIC	24499	11938	8281	6151	5523	5881
FMNB-3	AIC	24380	11787	8134	5995	5409	5776
	BIC	24738	12144	8492	6353	5767	6134

Notes:

AIC = $-2 \log(L) + 2K$; BIC = $-2 \log(L) + K \log(N)$; where L , K , and N are the maximized log likelihood, number of parameters, and observations, respectively.

^a model preferred by the AIC within the negative binomial- i class.

^b model preferred by the AIC overall.

^c model preferred by the BIC within the negative binomial- i class.

^d model preferred by the BIC overall.

either CFMNB-2 or FMNB-2 are preferred models. While there is some evidence in favour of NB1 on the basis of BIC for EMR and HOSP, evidence in favour of HNB is restricted to the case of OFP in the NB2 using AIC. To emphasize, neither HNB nor FMNB-3 are preferred to the FMNB-2 model. As expected, BIC favours the more parsimonious CFMNB-2 model whereas AIC favours the FMNB-2 model in three cases.

Taking together the evidence from the statistics described above we conclude that the FMNB-2 model is best within each density class and that models based on NB1 specifications perform better overall than those based on NB2 specifications. Thus, in what follows we present results from the FMNB-2 NB1 models. This evidence in favour of a two-component mixture also allows us to interpret the two populations as being healthy and ill.

Table V. Actual, fitted distributions and goodness-of-fit tests: FMNB-2 NB1 models

Count	<i>OFP</i>	<i>OFNP</i>	<i>OPP</i>	<i>OPNP</i>	<i>EMR</i>	<i>HOSP</i>
Sample frequencies						
0	15.5	68.2	77.1	83.8	81.8	80.4
1	10.9	12.9	11.9	9.8	13.3	13.6
2	9.7	4.7	4.6	2.8	3.1	4.0
3	9.5	2.8	1.7	1.1	1.2	1.1
4	8.7	2.4	1.2	0.6	0.2	0.5
5	7.7	1.6	0.8	0.2	0.2	0.3
6+	38.0	7.5	2.6	1.6	0.2	0.2
Fitted frequencies						
0	15.1	68.3	77.2	83.8	81.8	80.4
1	11.5	11.9	12.0	9.5	13.1	13.6
2	10.5	5.7	4.3	3.1	3.4	3.8
3	9.4	3.2	1.9	1.2	1.0	1.2
4	8.2	2.0	1.0	0.5	0.4	0.5
5	7.0	1.4	0.6	0.3	0.2	0.2
6+	38.3	7.5	2.9	1.6	0.2	0.3
Goodness-of-fit tests						
χ^2	14.80	25.25 ^a	13.69 ^a	3.13	4.46	0.59
df	13	8	5	2	2	2

Notes:

The fitted frequencies are the sample averages of the cell frequencies estimated from the FMNB-2 NB1 models.

^a Statistically significant at the 5% level.

6.2. Further Analysis of the Preferred Model

Table V presents the sample frequency distribution of the count variable along with the sample averages of the estimated cell frequencies from the selected model. We then present two views of the differences between the two populations that comprise the mixture. Table VI reports the fitted mean and variances for the fitted component densities and Figure 2 presents the fitted frequency distributions. Finally, Tables VII(a) and VII(b) contain parameter estimates for the FMNB-2 models estimated using NB1 specifications.

Tables VI, VII(a), and VII(b) report statistics for each of the two-component densities of the finite mixture. For office and hospital outpatient visits (*OFP*, *OFNP*, *OPP* and *OPNP*), the first component corresponds to the healthy population while the second component corresponds to the ill population. However, these interpretations are reversed for *EMR* and *HOSP*.

Goodness of fit

A comparison of sample and fitted frequency distributions in Table V shows a good fit over the entire range of the distribution. The discrepancy between the actual and fitted cell frequencies is never greater than 1%. Discrepancies between actual and predicted frequencies based on NB and HNB models (not presented) are usually much larger.

The χ^2 goodness-of-fit statistics were calculated for all models but are shown in Table V only for the selected models. The NB and HNB (both variants) are generally rejected by this test,

Table VI. Sample averages of fitted means and variances: FMNB-2 NB1 models

	<i>OFP</i>	<i>OFNP</i>	<i>OPP</i>	<i>OPNP</i>	<i>EMR</i>	<i>HOSP</i>
Mean						
Component density 1	5.55	0.58	0.33	0.39	0.35	0.99
Component density 2	8.17	2.43	2.11	0.67	0.22	0.26
Mixture density	5.78	1.62	0.76	0.54	0.26	0.30
Variance						
Component density 1	24.69	1.26	0.61	0.54	0.53	1.42
Component density 2	161.93	42.62	28.75	22.22	0.29	0.37
Mixture density	42.82	25.79	9.35	12.34	0.43	0.64

suggesting specification errors (not presented). The mixture models do considerably better in relative terms. In two cases (*OFNP* and *OPP*), however, the mixture specification is also rejected by this test—a surprising result given the match between sample and predicted cell frequencies—which reinforces the suspicion that the version of the test used here has a tendency to overreject a true null.

Estimates of π , component means, and densities

The estimates of the π components vary from a low value of 0.06 for *HOSP* to a high value of 0.91 for *OFP*. In all cases the estimates $\hat{\pi}$ and $(1 - \hat{\pi})$ are large relative to their estimated standard errors, reinforcing the conclusion that the evidence supporting the two population hypothesis is strong. Note, however, that rejecting the hypothesis of one population against the alternative of two populations, on the basis of the 't-test' of $\hat{\pi}$, has to be done with caution because the hypothesis involves a boundary value and the distribution of the test statistic is non-standard.

The sample averages of the estimates of $\lambda_{j,i}$ (fitted means) for the two component distributions are dramatically different in some cases. Healthy individuals who comprise 91% of the population have 5.6 visits to a physician (*OFP*) while the remaining ill individuals seek care 8.2 times. The component distributions shown in Figure 2 suggests that this difference in means is caused by a greater proportion of zeros and high values for the ill population. It appears that, while most healthy individuals see a doctor a few times a year for health maintenance and minor illnesses, a larger fraction of the ill do not seek any (preventive) care. Those ill individuals who do see a doctor (these individuals have sickness events) do so much more often than healthy individuals. Since preventive care usually takes the form of a visit to a doctor in an office setting, one would not expect such a pattern of differences between the healthy and ill to arise for the other measures of utilization.

The means are also quite different for *OFNP* and *OPPD*. While 44% of the population average 0.58 visits to a non-physician in an office (*OFNP*), the remainder average 2.43 visits. The mean number of visits to a physician in a hospital outpatient (*OPP*) setting is 0.33 for 76% of the population who are healthy but the remaining 24%, who are ill, seek care an average of 2.11 times. For these two cases, the component frequency distributions for *OFNP* and *OPP* (Figure 2) show that the main reason for the differences in the component means is the higher proportion of large counts in the ill population. There is little or no difference in the number of zero counts. This contrasts with the explanation for the differences in means for *OFP* where the healthy and ill groups differ in zero and large counts.

Table VII(a). Parameter estimates of FMNB-2 NB1 models

Variable	OFP		OFNP		OPP	
	Comp-1	Comp-2	Comp-1	Comp-2	Comp-1	Comp-2
<i>EXCLHLTH</i>	−0.25 ^a (0.06)	−0.77 (0.57)	−0.31 (0.35)	−0.08 (0.24)	−0.44 ^b (0.24)	−0.21 (0.48)
<i>POORHLTH</i>	0.24 ^a (0.07)	0.06 (0.89)	−0.52 (0.47)	−0.09 (0.20)	0.05 (0.19)	0.39 (0.34)
<i>NUMCHRON</i>	0.20 ^a (0.01)	0.14 (0.11)	0.14 ^b (0.09)	0.11 ^a (0.05)	0.16 ^a (0.04)	0.35 ^a (0.10)
<i>ADLDIFF</i>	0.01 (0.04)	0.58 (0.39)	−0.31 (0.34)	0.25 (0.16)	2e-3 (0.19)	0.30 (0.37)
<i>NOREAST</i>	0.08 (0.05)	0.21 (0.44)	2e-4 (0.41)	0.57 ^a (0.29)	−0.07 (0.18)	0.31 (0.38)
<i>MIDWEST</i>	0.01 (0.04)	0.09 (0.33)	−0.33 (0.35)	0.75 ^a (0.25)	0.11 (0.17)	0.24 (0.41)
<i>WEST</i>	0.09 ^b (0.05)	0.23 (0.44)	−0.10 (0.59)	1.01 ^a (0.23)	0.18 (0.17)	0.45 (0.33)
<i>AGE</i>	0.03 (0.02)	−0.57 ^a (0.20)	0.13 (0.47)	−0.17 (0.29)	−0.02 (0.10)	−0.40 ^a (0.18)
<i>BLACK</i>	−0.07 (0.06)	−1.16 (1.10)	−0.50 (0.52)	−0.19 (0.27)	−0.13 (0.35)	1.25 ^a (0.29)
<i>MALE</i>	−0.12 ^a (0.04)	0.06 (0.28)	−0.37 (0.27)	−0.28 ^b (0.16)	0.03 (0.12)	0.39 ^b (0.21)
<i>MARRIED</i>	0.04 (0.04)	−0.44 (0.38)	−0.04 (0.34)	0.13 (0.22)	0.05 (0.12)	0.16 (0.25)
<i>SCHOOL</i>	0.01 ^a (4e-3)	0.15 ^a (0.05)	0.05 (0.04)	0.06 ^a (0.03)	0.03 (0.02)	0.06 (0.04)
<i>FAMINC</i>	−3e-4 (0.01)	−1e-3 (0.01)	0.01 (0.01)	−0.02 (0.02)	−0.01 (0.02)	0.01 (0.03)
<i>EMPLOYED</i>	−0.06 (0.05)	0.39 (0.52)	0.28 (0.54)	−0.24 (0.33)	0.04 (0.20)	−0.70 (0.45)
<i>PRIVINS</i>	0.25 ^a (0.05)	2.89 (1.84)	0.68 ^b (0.40)	0.40 (0.27)	0.50 ^b (0.28)	−0.69 ^a (0.27)
<i>MEDICAID</i>	0.34 ^a (0.06)	−2.44 ^a (0.99)	−0.60 (0.94)	0.31 (0.28)	−0.02 (0.29)	−0.14 (0.35)
<i>CONSTANT</i>	0.78 ^a (0.18)	1.71 ^a (0.63)	−2.41 (3.77)	0.57 (2.47)	−1.97 ^a (0.71)	2.21 ^b (1.14)
α	3.45 ^a (0.19)	18.82 ^a (0.58)	1.18 (0.93)	16.51 ^a (0.68)	0.87 ^a (0.24)	12.63 ^a (3.25)
π_1	0.91 ^c (0.02)		0.44 ^c (0.21)		0.76 ^c (0.10)	
log <i>L</i>	−12072.8		−5836.2		−4021.3	

Notes:
Comp-1 and Comp-2 are the first and second component densities, respectively.
Standard errors are in parentheses.
^a Statistically significant at the 5% level.
^b Statistically significant at the 10% level.
^c Significantly different from 0 and 1 at the 5% level.

Table VII(b). Parameter estimates of FMNB-2 NB1 models

Variable	OPNP		EMR		HOSP	
	Comp-1	Comp-2	Comp-1	Comp-2	Comp-1	Comp-2
<i>EXCLHLTH</i>	-0.49 ^a (0.30)	-0.91 (0.93)	0.73 0.60	-4.36 ^a (0.82)	-3.26 (6.20)	-0.45 ^a (0.20)
<i>POORHLTH</i>	-0.22 (0.23)	0.61 (0.40)	1.31 ^a 0.64	-0.19 (0.28)	0.54 ^b (0.31)	0.54 ^a (0.10)
<i>NUMCHRON</i>	0.24 ^a (0.06)	0.09 (0.08)	0.36 ^a 0.06	0.13 ^a (0.06)	0.35 ^a (0.11)	0.27 ^a (0.03)
<i>ADLDIFF</i>	-0.12 (0.17)	0.83 ^a (0.28)	0.47 ^b 0.27	0.34 ^a (0.17)	0.85 ^a (0.33)	0.32 ^a (0.09)
<i>NOREAST</i>	0.50 ^a (0.16)	-0.08 (0.42)	-0.01 0.38	0.05 (0.24)	-4.97 ^a (2.25)	0.17 (0.13)
<i>MIDWEST</i>	0.22 (0.15)	0.52 ^a (0.25)	-0.21 0.32	0.09 (0.17)	0.51 (0.39)	0.04 (0.10)
<i>WEST</i>	0.33 (0.27)	-0.70 (0.93)	0.30 0.48	0.04 (0.22)	0.37 (0.30)	0.03 (0.11)
<i>AGE</i>	-0.24 ^a (0.09)	-0.61 ^a (0.18)	-0.42 ^a 0.19	0.36 ^a (0.08)	0.39 ^a (0.16)	0.13 ^a (0.06)
<i>BLACK</i>	-0.35 (0.24)	0.63 ^b (0.34)	0.50 0.34	-0.09 (0.38)	2.61 (1.90)	0.22 ^b (0.13)
<i>MALE</i>	0.31 ^a (0.13)	0.08 (0.27)	0.12 0.27	0.04 (0.18)	-0.04 (0.30)	0.22 ^a (0.08)
<i>MARRIED</i>	-0.01 (0.17)	-0.09 (0.29)	-0.02 (0.26)	-0.13 (0.15)	-0.19 (0.29)	-0.02 (0.09)
<i>SCHOOL</i>	0.04 (0.02)	0.04 (0.04)	-0.04 (0.03)	-4e-4 (0.02)	0.02 (0.05)	0.01 (0.01)
<i>FAMINC</i>	-3e-3 (0.02)	3e-3 (0.02)	-0.04 (0.09)	0.02 (0.02)	0.01 (0.04)	2e-3 (0.01)
<i>EMPLOYED</i>	-0.35 (0.30)	0.34 (0.44)	0.63 (0.46)	-0.15 (0.33)	-2.31 (1.73)	0.18 (0.14)
<i>PRIVINS</i>	0.68 ^a (0.18)	0.63 (0.39)	0.17 (0.29)	-4e-3 (0.14)	-0.05 (0.40)	0.13 (0.10)
<i>MEDICAID</i>	0.08 (0.23)	-0.33 (0.51)	-0.10 (0.43)	0.37 (0.26)	-3.97 ^a (1.87)	0.32 ^a (0.13)
<i>CONSTANT</i>	-0.49 (0.62)	2.61 ^a (1.21)	0.92 (0.58)	-4.46 ^a (0.68)	-3.94 ^a (1.51)	-3.35 ^a (0.47)
α	0.39 (0.38)	32.01 ^a (15.54)	0.51 ^b (0.28)	0.30 ^a (0.11)	0.44 ^b (0.26)	0.41 ^a (0.08)
π_1	0.46 ^c (0.21)		0.32 ^c (0.09)		0.06 ^c (0.02)	
	-2954.0		-2648.6		-2827.5	

Notes:

Comp-1 and Comp-2 are the first and second component densities, respectively.

Standard errors are in parentheses.

^a Statistically significant at the 5% level.^b Statistically significant at the 10% level.^c Significantly different from 0 and 1 at the 5% level.

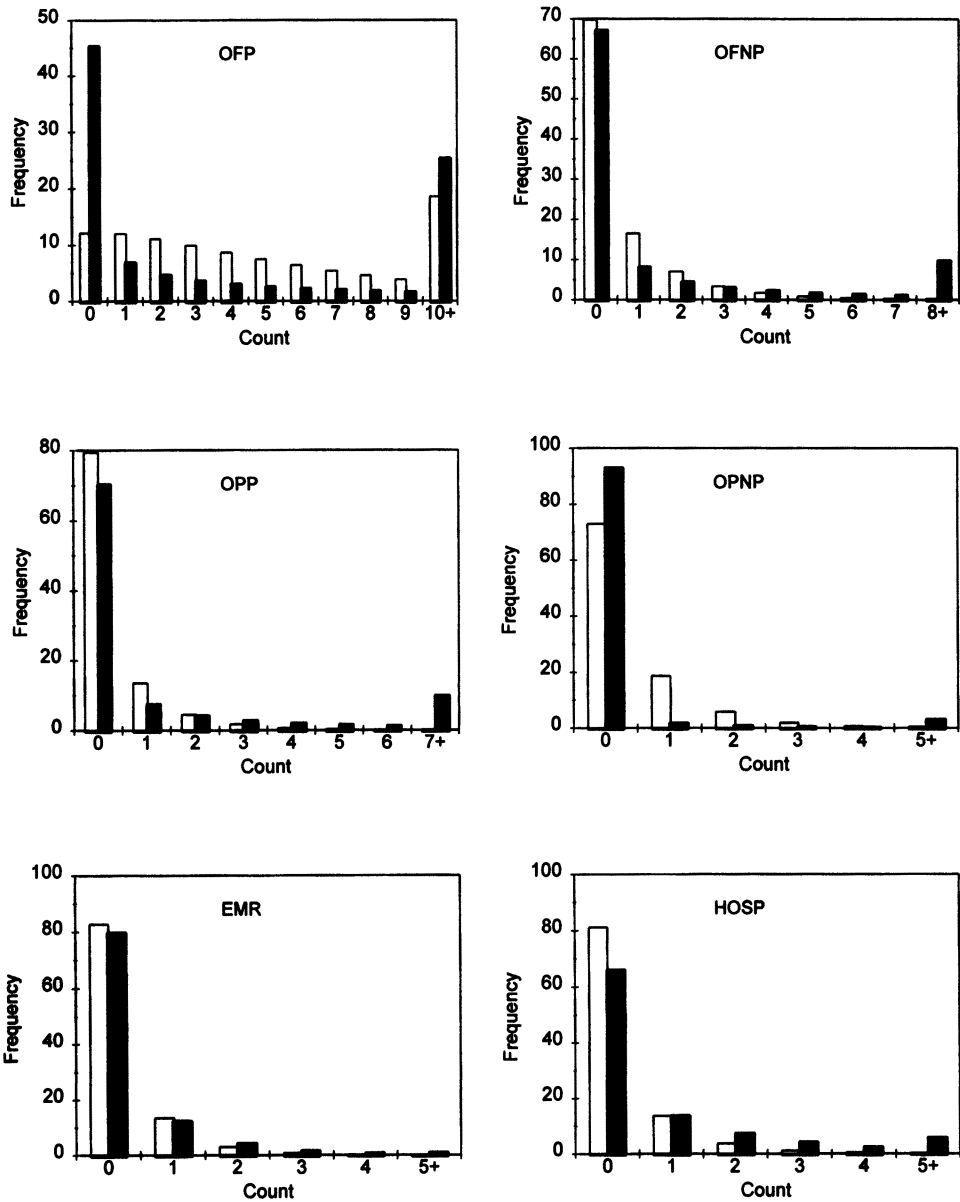


Figure 2. Component Densities from the FMNB-2 NB1 Model

The difference in means, though not large for the number of hospital stays (*HOSP*), is likely to be economically significant because of the higher costs of treatment in hospitals. As the distributions in Figure 2 show, about 19% of the healthy group, and 34% of the ill group, are likely to have at least one hospital stay. Finally, for the remaining two measures, *OPNP* and *EMR*, the component densities are not as markedly different as for other cases.

Estimates of α , variance, and overdispersion

The estimates of α are significantly different from zero, reinforcing the stylized fact that medical care utilization counts are overdispersed relative to the Poisson model.⁵ This is also apparent from the fact that the component means are generally smaller than the sample averages of the estimated component variances $\lambda_{j,i} + \alpha_j$ (Table VI). The sample average of the fitted variance of the mixture distribution is smaller than the sample variance of the corresponding count variable in each case. This is expected because the fitted variance accounts for unobserved heterogeneity while the sample variance accounts for observed heterogeneity (via covariates) and unobserved heterogeneity. The sample averages of the fitted variances for the two-component distributions are dramatically different in every case. Part of this difference is due to the fact that the variance of a negative binomial-1 variate is positively related to its mean. However, the estimates of α that correspond to the high mean component are much larger than those corresponding to the low mean component, suggesting that the ill population is more overdispersed than the healthy population.⁶ A large component of this difference comes from the significant differences in the probability of zero health events in the two populations. The histograms for event frequencies bring this out clearly. Overall, therefore, there is strong support for the hypothesis of two populations, one with a low mean and variance, and the other with a higher mean and higher variance.

Insurance, income, and employment

Supplementary private insurance status (*PRIVINS*) does not affect emergency room or hospital utilization because Medicare coverage is quite generous in these areas (the net price to the user is quite small, and often zero). Utilization is insensitive to marginal changes in price given the small baseline price levels. Persons with supplementary private insurance seek care from physicians and non-physicians in office as well as outpatient settings more often than individuals without supplementary coverage and the additional usage for *OFP* is estimated to be considerably higher for the high-use group. This is because private insurance typically covers physical therapy, check-ups, etc., with small deductibles and coinsurance rates while Medicare does not.⁷

Medicaid coverage is a significant determinant of the number of visits for two categories of medical care—*OFP* and *HOSP*—which are the two most important measures of utilization: office visits to a physician in terms of the overall number of medical care visits and hospital stays in terms of the share of expenditures on medical care. In both cases, the coefficient is significantly positive in the component density for the healthy group and is significantly negative in the component density for the ill group. A plausible explanation for this intriguing result is that, for the healthy group, the price (insurance) effect of the *MEDICAID* dummy outweighs the income effect (because *MEDICAID* is a health insurance plan for the poor). But poor individuals within the ill group (for whom the opportunity cost of seeking care is disproportionately large relative to the money-price of care) seek care less often even though they have Medicaid insurance coverage. This is a potentially important finding, and one that is uncovered in the finite mixture framework in a natural way, which confirms that the unhealthy poor are a vulnerable population in society.

⁵ The *t*-test is not strictly valid because α is bounded below by zero.

⁶ The parameter estimate of α for the second component of OPNP is implausibly large (32.01). Therefore, we re-estimated the model with a limited number of different starting values but they yielded little change in the maximized estimates. Note, however, that our discomfort with this value is statistically embodied in the large estimated standard error (15.54).

⁷ Because we do not have data on the price of care, however, we cannot estimate price elasticities of demand, which would be of practical use.

Perhaps surprisingly, the (continuous) family income variable (*FAMINC*) does not affect the utilization of medical care. Furthermore, employment, which may capture income effects as well, is never significant. One explanation for the negligible income effect is that the overall generosity of Medicare irrespective of family income, combined with Social Security Income which guarantees a certain level of income, make utilization insensitive to marginal changes in income.⁸ Another is that the effect of income is to induce changes in the intensity and quality of care which would be captured in expenditure functions but not in visits functions.

Health status

Consistent with previous studies, both the number of chronic conditions (*NUMCHRON*) and self-perceived health are important determinants of utilization. An increase in the number of chronic conditions increases utilization of each form of medical care.⁹ The measures of self-perceived health (*EXCLHLTH* and *POORHLTH*) affect *OFP* but not *OFNP* or *OPP*. *HOSP* and *EMR* are increased by perceptions of ill-health. Disability (*ADLDIFF*) increases emergency room and hospital care usage but generally does not affect other forms of care. It is well known that emergency room and inpatient care are the most expensive forms of care. Moreover, the debilitating impact of a number of conditions that lead to disability can be ameliorated or postponed through physical therapy or rehabilitation. Since physical therapy and some rehabilitative programmes are considerably cheaper than hospital stays, a policy implication of this finding is that resources could be spent on programmes that reduce limitations.

Demographics

Hospital stays increase with age but hospital outpatient visits decrease with age, probably due to decreasing mobility associated with ageing.¹⁰ Men seek less care in office settings and from non-physicians in outpatient settings than women but have more hospital stays and outpatient physician visits. An explanation for this phenomenon is the anecdotal 'fact' that men tend to wait longer before seeking medical care, and therefore have more serious conditions that require treatment in in- and outpatient hospital settings. Cartwright, Hu, and Huang (1992) corroborate these findings and explanation; they find that men are less likely to seek medical care, but once care is sought they expend about 13% more than women.

Individuals with greater schooling seek care in office settings more often than those with less. Given the insignificance of income, this finding could be due to an income effect signalled through schooling, but it could also be due to the more conventional reason that education makes individuals more informed consumers of medical care services. Marital status appears to have no effect on utilization. Finally, the region of residence appears to have some influence on the mix of office and outpatient utilization but not on emergency room or hospital visits.

7. CONCLUSION

The results obtained from the three variants of the negative binomial count-regression (standard, hurdle, and finite mixture) estimated in this article lend support to the theoretical models of

⁸ This finding is consistent with the literature; see, for example, Cartwright, Hu, and Huang (1992).

⁹ We present results using one variable measuring the number of chronic conditions for parsimony. However, the conclusion with respect to the effect of chronic conditions on utilization is not altered when a dummy variable is used for each condition.

¹⁰ Doctors' offices which are located in diverse geographic locations are typically more accessible than outpatient clinics in large hospitals, which are concentrated in city centres.

medical care demand in the consumer-theoretic framework which do not need to assume that the decision making for the demand for health care has a two-part structure. To the extent that a two-part (hurdles) specification is interpreted as reflecting a principal–agent structure, our empirical results do not support it. However, it should be acknowledged that a hurdles model need not be interpreted as a unique characterization of the principal–agent type formulation. With the exception of the hospitalization outcome, our results support a two-point finite mixture model, thus allowing a categorization of individuals as ill and healthy, a characterization which is silent on how the patient–provider relationship leads to a particular outcome.

Our empirical results emphasize the importance of health status and insurance as determinants of health-care demand, and the relative unimportance of income.

An interesting issue for future research concerns the use of a multivariate framework which will permit one to consider whether ‘low user’ and ‘high users’ categories across different types of usage are significantly and positively correlated, as indicated by crude correlation measures based on raw data. If so, this will lend additional support to such a dichotomy, which in turn is germane to the distinction between ‘high risks’ and ‘low risks’ in models of adverse selection in health insurance. A second methodological issue for future research is the avoidance of estimators based on strong distributional assumptions in favour of moment-based estimators.

ACKNOWLEDGEMENTS

We thank three referees and Kurt Brännäs for their helpful comments which helped to improve earlier drafts of the paper. Helpful comments were also obtained from workshop presentations at Tilburg and Munich. All remaining errors are ours.

REFERENCES

- Aitken, M. and D. B. Rubin (1985), ‘Estimation and hypothesis testing in finite mixture models’, *Journal of the Royal Statistical Society, Series B*, **47**, 67–75.
- Andrews, D. W. K. (1988), ‘Chi-square diagnostic tests for econometric models’, *Journal of Econometrics*, **37**, 135–56.
- Böhning, D. (1995), ‘A review of reliable maximum likelihood algorithms for semiparametric mixture models’, *Journal of Statistical Planning and Inference*, **47**, 5–28.
- Böhning, D., E. Dietz, R. Schaub, P. Schlattmann, and B. G. Lindsay (1994), ‘The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family’, *Annals of the Institute of Statistical Mathematics*, **46**, 2, 373–88.
- Brännäs, K. and G. Rosenqvist (1994), ‘Semiparametric models of heterogeneous count data models’, *European Journal of Operations Research*, **76**, 247–58.
- Cameron, C. and P. K. Trivedi (1986), ‘Econometric models based on count data: comparisons and applications of some estimators and tests’, *Journal of Applied Econometrics*, **1**, 29–53.
- Cameron, C., P. K. Trivedi, F. Milne, and J. Piggot (1988), ‘A microeconomic model of the demand for health care and health insurance in Australia’, *Review of Economic Studies*, **55**, 85–106.
- Cartwright, W. S., T.-W. Hu, and L.-F. Huang (1992), ‘Impact of varying Medigap insurance coverage on the use of medical services of the elderly’, *Applied Economics*, **24**, 529–39.
- Christensen, S., S. Long, and J. Rodgers (1987), ‘Acute health care costs for the aged Medicare population: overview and policy options’, *Millbank Quarterly*, **65**, 397–425.
- Duan, N., W. Manning, C. Morris, and J. Newhouse (1983), ‘A comparison of alternative models for the demand for medical care’, *Journal of Business and Economic Statistics*, **1**, 115–26.
- El-Gamal, M. and D. Grether (1995), ‘Are people Bayesian? Uncovering behavioral strategies’, *Journal of the American Statistical Association*, **90**, 1137–45.

- Grossman, M. (1972), 'On the concept of health capital and the demand for health', *Journal of Political Economy*, **80**, 223–55.
- Gurmu, S. and P. K. Trivedi (1996), 'Excess zeros in count models of recreational trips', *Journal of Business and Economic Statistics*, **14**, 4, 469–77.
- Gurmu, S. (1997), 'Semiparametric estimation of hurdle regression models with an application to Medicaid utilization', *Journal of Applied Econometrics*, **12**, 225–242.
- Heckman, J. J. (1984), 'The χ^2 goodness of fit statistic for models with parameters estimated from microdata', *Econometrica*, **52**, 1543–47.
- Heckman, J. J. and B. Singer (1984), 'A method of minimizing the distributional impact in econometric models for duration data', *Econometrica*, **52**, 271–320.
- Huang, L.-F. and O. Koropecy (1976), 'An analysis of the demand response to complementary insurance under Medicare for the aged', *Journal of the American Statistical Association: Proceedings, Business and Economics Section*.
- Johnson, N. L., S. Kotz and A. W. Kemp (1992), *Univariate Distributions*, 2nd edition, Wiley, New York.
- Laird, N. (1978), 'Nonparametric maximum likelihood estimation of a mixing distribution', *Journal of the American Statistical Association*, **73**, 805–11.
- Leroux, B. G. (1992), 'Consistent estimation of a mixing distribution', *Annals of Statistics*, **20**, 3, 1350–60.
- Lindsay, B. J. (1995), *Mixture Models: Theory, Geometry, and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5, IMS-ASA.
- Lindsay, B. J. and K. Roeder (1992), 'Residual diagnostics in the mixture model', *Journal of the American Statistical Association*, **87**, 785–95.
- McLachlan, G. J. and K. E. Basford (1988), *Mixture Models: Inference and Application to Clustering*, Marcel Dekker, New York.
- Manning, W. G., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz, and M. Marquis (1987), 'Health insurance and the demand for medical care: evidence from a randomized experiment', *American Economic Review*, **77**, 251–77.
- Muurinen, J.-M. (1982), 'Demand for health, a generalized Grossman model', *Journal of Health Economics*, **1**, 3–28.
- Pohlmeier, W. and V. Ulrich (1995), 'An econometric model of the two-part decisionmaking process in the demand for health care', *The Journal of Human Resources*, **XXX**, 339–61.
- Rice, T. (1984), 'Physician induced demand for medical care: new evidence from the Medicare program', *Advances in Health Economics and Health Services Research*, **5**, 129–60.
- Waldo, D. R., S. T. Sonnefeld, D. R. McKusick, and R. H. Arnett, III (1989), 'Health expenditures by age group, 1977 and 1987', *Health Care Financing Review*, **10**, 111–20.
- Wedel, M., W. S. Desarbo, J. R. Bult, and V. Ramaswamy (1993), 'A latent class Poisson regression model for heterogeneous count data', *Journal of Applied Econometrics*, **8**, 397–411.
- Zweifel, P. (1981), 'Supplier-induced demand in a model of physician behavior', in J. van der Gaag and M. Perlman (eds), *Health, Economics, and Health Economics* 245–67, North-Holland, Amsterdam.