Likelihood-Based Approaches to Modeling Demand for Medical Care

Michael Creel and Montserrat Farell*

October 2001

Abstract

We review recent likelihood-based approaches to modeling demand for medical care. A semi-nonparametric model along the lines of Cameron and Johansson's Poisson polynomial model, but using a negative binomial baseline model, is introduced. We apply these models, as well a semi-parametric Poisson, hurdle semiparametric Poisson, and finite mixtures of negative binomial models to six measures of health care usage taken from the Medical Expenditure Panel survey. We conclude that most of the models lead to statistically similar results, both in terms of information criteria and conditional and unconditional prediction. This suggests that applied researchers may not need to be overly concerned with the choice of which of these models they use to analyze data on health care demand.

Keywords: health care demand, count data, maximum likelihood

JEL classifications: C25, I10

*Both authors are at the Department of Economics and Economic History, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) Spain. Send email to michael.creel@uab.es. This research was funded by DGICYT grant SEC 2000-0390.

1 Introduction

Demand for medical care - doctor visits, prescription drugs taken, *etc.*, is usually observed as an event count over some time period. When possible, it is desirable for efficiency reasons to estimate by maximum likelihood, which implies the use of count data models. This paper offers a review of recent likelihood-based methods for explaining demand for health care, when demand is observed as a count. The models are applied to the 1996 Medical Expenditure Panel Survey data, for six measures of health care usage. With few exceptions, the different statistical models achieve very similar values of information criteria and in- and out-of-sample fit. Looking at predicted usage conditional on income and age, we again observe few differences between most of the models. The important conclusion of the paper is that which among the models considered is used for a policy analysis is unlikely to importantly affect the conclusions of the analysis. This suggests that applied researchers might more profitably concentrate on issues such as the selection of conditioning variables or the parameterization of the model, rather than the choice of the density function. The densities employed in this paper appear to be sufficiently flexible to capture the important features of the sort of data under consideration.

Health care demand often exhibits overdispersion, in that the ratio of the conditional variance to the conditional mean is greater than one (Cameron and Trivedi, 1986; Pohlmeier and Uhlrich, 1995). Another common characteristic is that many zeros are observed, more than can be accounted for by simple count densities (Pohlmeier and Uhlrich, 1995; Gerdtham, 1997). To better focus on the special characteristics of this sort of data, and to relate the econometric methods to theoretical issues in the analysis of the demand for medical care, we limit our attention to the demand for medical care, though the count data models that are considered here could be applied in areas such as analysis of the production of patents or demand for recreation trips.

One important issue from which we will abstract is that of the exogeneity of the explanatory variables. Variables such as private insurance coverage or self-reported health status may be jointly determined with variables related to usage of health care services (Cameron *et. al.*, 1988; Windmeijer and Santos Silva, 1997; Vera-Hernández, 1999). For example, if both usage and the decision to purchase private insurance are in part determined by an unobservable personal characteristic such as health status, then there will exist a problem of endogeneity in the estimation of the usage model, if usage depends upon insurance status. When this problem is encountered, a possibility is to step away from distributional assumptions and to estimate using GMM. A disadvantage of this approach is that GMM estimates, especially for the coefficients of the suspected endogenous variables, may be quite imprecise (Windmeijer and Santos Silva, 1999; Vera-Hernández, 1999). Another possibility is to estimate a joint density for all endogenous variables using MLE (Cameron and Johansson, 1998; Romeu and Vera-Hernández, 2001), though this has not been completely developed. Our primary goal in this paper is to compare the ability of likelihood-based approaches to fit data on demand for medical care, supposing that the regressors may be treated as exogenous. The papers by Windmeijer and Santos Silva (1997) and Vera-Hernández (1999) find mixed evidence regarding endogeneity of self reported health status, one the one hand, and duplicate insurance coverage, on the other. Hausman tests do not reject exogeneity when based only on the coefficient of the possibly endogenous variable, but they do reject when based upon all coefficients. It seems that a conclusive answer will depend upon formulating a GMM estimator using more informative moment conditions, or perhaps using the joint density approach for all endogenous variables. While we expect that problems of endogeneity will sometimes complicate or rule out single equation MLE for health care usage as a method for obtaining consistent estimates, we nevertheless believe that even in these cases it is an empirically interesting question to know which models can fit best, since this knowledge might assist in

formulating moment conditions upon which to base a GMM estimation strategy. In the case that the regressors can be treated as exogeneous, a comparision of the statistical models upon which MLE is based is clearly relevant in its own right.

Before moving on to compare the statistical models, we briefly note how theoretical perspectives regarding the demand for medical care have been reflected in the empirical literature. Grossman's seminal paper (1972) offers a human capital perspective that treats health as a capital good that is subject to depreciation. Health care services are inputs to the production of health capital. This model treats the individual as the prime decision maker regarding the consumption of health care services. Cameron *et. al.* (1988) present a model that is within this tradition. This paper is also among the first to emphasize the count data aspect of health care visits. Zweifel (1981) presents a principal-agent model that recognizes that while the individual may initiate contact with the physician, as in Grossman's model, the physician will have much weight in deciding the treatment, which will be important in determining the number of follow-up visits. This idea is incorporated in empirical work by Manning *et. al.* (1987), who do not address count data aspects. Pohlmeier and Ulrich (1995) and Gerdtham (1997) are examples of papers that build on the principle-agent perspective and also account for count data issues. The more recent papers mentioned below incorporate additional statistical refinements to these perspectives.

2 A survey of recent approaches

In this section we briefly survey some of the newer count data models that have been proposed for demand for health care services. Before surveying the recent models, we briefly discuss the more standard models upon which the newer approaches build.

Poisson

The Poisson density for a count random variable Y is

$$f_Y(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}.$$

To allow for covariates, λ is usually parameterized as $\lambda = e^{\mathbf{x}\beta}$. The Poisson density implies that the conditional mean and the conditional variance of y are both equal to λ. Since data on health care demand usually exhibit overdispersion and excess zeros, the basic Poisson model will usually not be suitable for analyzing demand for health care.

Negative binomial (NB)

If the Poisson mean contains a latent component, marginalization, under some assumptions, will lead to a negative binomial density . The negative binomial density may be written as

$$f_Y(y|\phi) = \frac{\Gamma(y+\psi)}{\Gamma(y+1)\Gamma(\psi)} \left(\frac{\psi}{\psi+\lambda}\right)^{\psi} \left(\frac{\lambda}{\psi+\lambda}\right)^{y} \tag{1}$$

where $\phi = \{\lambda, \psi\}$, $\lambda > 0$ and $\psi > 0$.[1] As with the Poisson models, the usual means of incorporating conditioning variables is the parameterization $\lambda = e^{\mathbf{x}\beta}$. When $\psi = \lambda/\alpha$ we have the negative binomial-I model (NB-I). When $\psi = 1/\alpha$ we have the negative binomial-II (NP-II) model. Though other versions exist, we limit attention to these in this paper. The moment generating function of the NB density, which is needed below, is

$$M_Y(t) = \psi^{\psi}\left(\lambda - e^t\lambda + \psi\right)^{-\psi}. \tag{2}$$

[1] Among the numerous examples of application of the NB model to health care demand are Cameron *et. al.* (1988) (1986), Pohlmeier and Ulrich (1995) and Geil *et. al.* (1997).

For the NB-I density, $V(Y) = \lambda + \alpha\lambda$. In the case of the NB-II model, we have $V(Y) = \lambda + \alpha\lambda^2$. For both forms, $E(Y) = \lambda$. Thus, both forms capture overdispersion, with the NB-II model allowing for a more extreme form. Nevertheless, as discussed in the introductory section, health care demand often exhibits excess zeros with respect to what a NB model can accomodate. This leads us to consider the hurdle version of the NB model.

Hurdle negative binomial (HNB)

The hurdle[2] model can be rationalized as a statistical representation of a principle/agent model, where individuals decide whether or not to seek care, but once care is sought, the physician influences how many visits will take place. It is natural to assume that different parameters govern the decisions of the principle and the agent. To accomodate this, the HNB model assumes that a individuals make a binary discrete choice of whether or not care is sought. This is modeled as a Bernoulli trial using a probit or similar model. Conditional on positive visits, the count follows a zero-truncated negative binomial (TNB) density. Hurdle count models were introduced by Cragg (1971) and Mullahy (1986), who also presented "with-zeros"[3] models. Here we present only the hurdle model, due to its more intuitive foundations relative to the zero inflated model, at least for the case of health care demand, and due to the fact that it seems to have been used more widely to model health care demand.[4] We follow Deb and Trivedi (1997), who use a NB model to parameterize the Bernoulli trial. For a NB random variable,

$$\begin{aligned} \Pr(Y=0) &= f_Y(0,\phi_h) = \left(\frac{\psi_h}{\psi_h+\lambda_h}\right)^{\psi_h} \\ \Pr(Y>0) &= 1-\Pr(Y=0), \end{aligned}$$

[2]Hurdle models are also known as "two-part" models.

[3]The "with-zeros" model is also known as the "zero-inflated" model.

[4]Examples of applications of the HNB model to health care demand include Pohlmeier and Ulrich (1995), Gerdtham (1997) and Deb and Trivedi (1997).

where the parameter of the hurdle process is $\phi_h = \{\lambda_h, \psi_h\}$. To achieve identification one can set $\alpha_h = 1$ when parameterizing ψ_h according to the NB-I or NB-II models. The above probabilities are used to estimate the binary 0/1 hurdle process. Then, for the observations where visits are positive, a TNB density, with a different parameter $\phi = \{\lambda, \phi\}$ is estimated. This density is

$$\begin{aligned} f_Y(y, \phi | y > 0) &= \frac{f_Y(y, \phi)}{1 - \left(\frac{\psi}{\psi+\lambda}\right)^{\psi}} \\ &= \frac{\Gamma(y+\psi)}{\Gamma(y+1)\Gamma(\psi)} \left[\left(\frac{\psi}{\psi+\lambda}\right)^{\psi} - 1\right]^{-1} \left(\frac{\lambda}{\psi+\lambda}\right)^{y} \end{aligned}$$

Since the hurdle and truncated components of the overall density for Y share no parameters, they may be estimated separately, which is computationally more efficient than estimating the overall model. The expectation of Y is

$$E(Y) = \left[1 - \left(\frac{\psi_h}{\psi_h + \lambda_h}\right)^{\psi_h}\right] \left[1 - \left(\frac{\psi}{\psi+\lambda}\right)^{\psi}\right]^{-1} \lambda$$

NB-I and NB-II versions follow from the appropriate parameterizations of ψ_h, ψ, λ_h and λ.

The HNB model can probably be considered the state of the art for modeling health care demand count data, up until 1997. Shortly after, the following models were introduced. All of these models can account for excess zeros and overdispersion, and they are more flexible than the HNB model, so that distributional assumptions are considerably relaxed.

2.1 A semiparametric approach (PSP, HPSP)

The semiparametric approach to modeling count data has been developed by Gurmu and Trivedi (1996), Gurmu (1997) and Gurmu *et. al.* (1999). This approach introduces unobserved heterogeneity in a Poisson model, and allows the unobserved heterogeneity to follow a semi-nonparametric density. The semi-nonparametric density is closely related to that proposed by Gallant and Nychka

(1987). The difference is that Laguerre polynomials are used instead of Hermite polynomials. Gurmu *et. al.* show that, under weak assumptions, the Laguerre expansion density can consistently estimate densities of unknown form. As such the mixture density is semiparametric, since the Poisson specification is parametric but the modelization of the heterogeneity is not.

Gurmu and Trivedi (1996) found that the basic semiparametric approach of Gurmu *et. al.* (1999)[5] did not fit data well. This is probably due to the reliance on the Poisson baseline model, which often does not fit well. To overcome this problem, Gurmu (1997) proposed a hurdle version of the semiparametric model.

The original semiparametric model is based upon an infinite mixture of a Poisson random variable and an independent random variable V which captures unobserved heterogeneity. The assumption is that the Poisson mean is random, so that $E(Y|V = v) = \lambda v$. Integrating out the heterogeneity, one obtains the marginal density:

$$f_Y(y, \lambda, \phi) = \int \frac{e^{-\lambda v}(\lambda v)^y}{y!} g_V(v, \phi) dv \tag{3}$$

$$= \frac{\lambda^y}{y!} M_V^y(-\lambda) \tag{4}$$

where $M_V^y(-\lambda)$ is the yth order derivative of the moment generating function of V, evaluated at $-\lambda$. $M_V^0(-\lambda) = M_V(-\lambda)$, is the moment generating function itself.

To model the density $g_V(v, \phi)$ flexibly, Gurmu *et. al.* use a normalized Laguerre polynomial expansion around a gamma baseline density. The gamma baseline density is

$$f(v, \phi) = \left(\frac{v^{\alpha-1}\beta^\alpha}{\Gamma(\alpha)} e^{-\beta v} \right)$$

[5]The 1999 paper is based upon a 1996 working paper, which explains the dates of these references.

where $\phi = (\alpha, \beta)$. The semi-nonparametric density for v is

$$g_V(v|\phi,\gamma) = \frac{[h_p(y,\gamma)]^2 f(v|\phi)}{\eta_p(\phi,\gamma)}$$

where

$$h_p(y,\gamma) = \sum_{k=0}^{p} \gamma_k P_k(v), \tag{5}$$

$\gamma = (1, \gamma_1, \gamma_2, ..., \gamma_p)$, and $P_k(v)$ is the kth order Laguerre polynomial. The term $\eta_p(\phi,\gamma) = \gamma\prime\gamma$ is the normalization factor that makes the density sum to one. The restriction that $\gamma_0 = 1$ is used to achieve identification, since the density homogeneous in γ. This density is semi-nonparametric in the sense that, under weak assumptions, there exist ϕ, γ such that a density of unknown form can be approximated arbitrarily well as p goes to infinity. Gurmu *et. al.* (1999) provide the consistency proof, which is closely modeled on that of Gallant and Nychka (1987).

Next, they are able to obtain a closed form for $M_V^y(-\lambda)$, which upon substitution into equation 4 yields the semiparametric density for the count random variable Y. In estimation, a restriction is imposed such that $E(V) = 1$, which leads to $E(Y) = \lambda$. In the course of the empirical work reported below, we have found that the model is poorly identified without this restriction, and that it is very difficult to obtain convergence if it is not imposed. The results we report alway impose the restriction. We will refer to this model as the Poisson semiparametric model (PSP).

To extend this to the hurdle case, Gurmu (1997) allows a first PSP model to determine whether the zero/positive hurdle is crossed, and a second PSP model is used to model the positives. For the hurdle crossing process, the relevant probabilities are

$$\begin{aligned} \Pr(Y=0) &= M_V(-\lambda_h) \\ \Pr(Y>0) &= 1 - \Pr(Y=0). \end{aligned}$$

The truncated version of the PSP density is

$$f_Y(y|y>0,\lambda,\phi) = \frac{\frac{\lambda^y}{y!}M_V^y(-\lambda)}{1-M_V(-\lambda)}.$$

Just as in the case of the HNB model, the binary and truncated components of the hurdle Poisson semiparametric (HPSP) model may be estimated separately.

2.2 Semi-nonparametric approaches (PSNP, NBSNP)

Cameron and Johansson (1997) directly adapt Gallant and Nychka's (1987) semi-nonparametric density to the count data case. They reshape a Poisson baseline density using a squared polynomial, and then normalize the result to sum to one. We shall refer to this as the Poisson semi-nonparametric (PSNP) approach, though there has been no formal proof of the conditions under which the density has nonparametric properties.[6] The PSP and HPSP models embed the semi-nonparametric density in a parametric density to obtain a semiparametric model, after marginalization of the latent variable. As such, one expects that the approach of Cameron and Johansson should be able to capture more extreme departures from the baseline model, though perhaps at the cost of needing to estimate many parameters. For example, the PSNP model can accomodate bimodal densities, while the PSP density cannot.

The PSNP density is

$$f_Y(y|\lambda,\gamma) = \frac{[h_p(y|\gamma)]^2}{\eta_p(\phi,\gamma)}\frac{e^{-\lambda}\lambda^y}{y!},$$

where

$$h_p(y|\gamma) = \sum_{k=0}^{p}\gamma_k y^k, \tag{6}$$

and $\eta_p(\phi,\gamma)$ is a normalizing factor to make the density sum to one. The

[6]The consistency proofs of Gallant and Nychka (1987) and Gurmu *et. al.* (1999) are for continuous random variables. While it seems reasonable to expect that the proofs could be adapted to discrete random variables, this has not yet been done, to our knowledge.

normalizing factor is

$$\eta_p(\lambda,\gamma) = \sum_{y=0}^{\infty} \left[h_p\left(y|\gamma\right)\right]^2 \frac{e^{-\lambda}\lambda^y}{y!}.$$

Cameron and Johansson show that this has the closed form

$$\eta_p(\lambda,\gamma) = \sum_{k=0}^{p}\sum_{l=0}^{p} \gamma_k\gamma_l m_{k+l} \tag{7}$$

where $m_r(\lambda)$ is the rth noncentral moment of the Poisson density. Because $\left[h_p\left(y|\gamma\right)\right]^2/\eta_p(\lambda,\gamma)$ is a homogenous function of γ, it is necessary to impose a normalization to achieve identification: γ_0 is set to 1. The moments of Y may be calculated using the closed form expression in Cameron and Johansson's equation 4.[7]

Since the NB model usually fits health care data dramatically better than does the Poisson model, using only one more parameter, one might suspect that changing the baseline model to the NB might allow the model to fit well with fewer parameters. What we shall refer to as the negative binomial semi-nonparametric (NBSNP) model is obtained by making this change. The density is

$$f_Y(y|\lambda,\gamma) = \frac{\left[h_p\left(y|\gamma\right)\right]^2}{\eta_p(\phi,\gamma)} \frac{\Gamma(y+\psi)}{\Gamma(y+1)\Gamma(\psi)} \left(\frac{\psi}{\psi+\lambda}\right)^{\psi} \left(\frac{\lambda}{\psi+\lambda}\right)^{y},$$

where $h_p\left(y|\gamma\right)$ and $\eta_p(\phi,\gamma)$ are defined as in equations 6 and 7, respectively, and the raw moments $m_r(\lambda,\psi)$ are obtained from equation 2. The moments of Y are again obtained from Cameron and Johannson's equation 4, after substituting the NB raw moments.[8]

[7] Cameron and Johansson (1998) have extended the PSNP approach to consider multivariate count data models.

[8] We used *Mathematica* (R) version 4.0 to perform these calculations.

2.3 A finite mixture approach (MNB, CMNB)

The finite mixture approach to fitting health care demand was introduced by Deb and Trivedi (1997). The mixture approach has the intuitive appeal of allowing for subgroups of the population with different health status. If there are two subgroups it is natural to think of healthy and unhealthy persons. Alternatively, one could argue for the existence of gravely ill persons, unhealthy persons, and healthy persons. Many studies have incorporated objective and/or subjective indicators of health status in an effort to capture this heterogeneity. Objective measures, such as limitations on activity, are not necessarily very informative about a person's overall health status. Subjective, self-reported measures may suffer from the same problem, and may also not be exogenous, in which case there are additional modeling issues (Windmeijer and Santos Silva, 1997). The finite mixture approach allows health status to be truly latent. The mixture approach has since been applied by Gerdtham and Trivedi (2000), who find that it performs better than the HNB approach.

The mixture negative binomial (MNB) model has the virtue of being conceptually simple. The density is

$$f_Y(y, \phi_1, ..., \phi_p, \pi_1, ..., \pi_{p-1}) = \sum_{i=1}^{p-1} \pi_i f_Y^{(i)}(y, \phi_i) + \pi_p f_Y^p(y, \phi_p),$$

where $\pi_i > 0, i = 1, 2, ..., p$, $\pi_p = 1 - \sum_{i=1}^{p-1} \pi_i$, and $\sum_{i=1}^{p} \pi_i = 1$. The $f_Y^{(i)}(y, \phi_i)$, $\phi_i = \{\lambda_i, \psi_i\}$ are p separate NB-I or NB-II densities, as in equation 1. Identification requires that the π_i are ordered in some way. We follow Deb and Trivedi (1997) by imposing $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_p$ and $\phi_i \neq \phi_j, i \neq j$. This is simple to accomplish post-estimation by rearrangement and possible elimination of redundant component densities.

The properties of the mixture density follow in a straightforward way from those of the components. In particular, the moment generating function is the same mixture of the moment generating functions of the component densities,

whence $E(Y) = \sum_{i=1}^{p} \pi_i \lambda_i$.

The MNB density may suffer from overparameterization, since the total number of parameters grows rapidly with the number of component densities. To address this problem, Deb and Trivedi propose a constrained mixture negative binomial model (CMNB) which restricts all the slope parameters in $\lambda_j = e^{\mathbf{x}\beta_j}$ to be the same across all component densities. The constants and the overdispersion parameters α_j are allowed to differ.

3 Data

We applied the above models to the 1996 Medical Expenditure Panel Survey data, using six different measures of annual health care usage[9]. These are office-based doctor visits (OBDV), outpatient visits (OPV), emergency room visits (ERV), inpatient visits (IPV), dental visits (DV), and number of prescription drugs taken (PRESCR). We limited the analysis to individuals between the ages of 18 and 65, inclusive. Women who experienced a pregnancy during 1996 were excluded from the sample. The explanatory variables are months of private insurance coverage during the year (PRIV), months of public insurance coverage during the year (PUB), sex (SEX), age (AGE), years of schooling (EDUC), and family income (INC). All the variables with the exception of INC are directly available. INC was constructed by summing the incomes of all members of the family. Observations for which any family member's income was "hot decked" were dropped.[10]

The number of observations for which this set of variables is complete is 4566. Because some of the models presented above can become numerically unstable when the dependent variable takes on large values, we dropped obser-

[9]The raw data is available at www.meps.ahrq.gov/MEPSDATA/HC/H12DAT.zip, and the programs used to prepare the data, the prepared data, and the estimation routines are available at pareto.uab.es/mcreel. All programming was done using Ox version 3.0 (Doornik, 1999).

[10]"Hot decking" is a term used in the MEPS documentation to describe a method of replacing missing data with conditional or unconditional means of the variable. See the documentation at www.meps.ahrq.gov/Pubdoc/H12DOC.PDF for more details.

vations when the respective dependent variable took on a value greater than 75. This was done for all models, to facilitate comparison. This implied dropping 4 observations for OBDV, 1 observation for OPV, and 80 observations for PRESCR. The other measures of health care usage were unaffected. Next, the sample was randomly divided into two portions, one for estimation and one for out-of-sample prediction. This was done by selecting observations for estimation or prediction according to whether the value of a uniform $[0, 1]$ random variable was less or greater than 0.5, so that the split is approximately 50-50.

Table I provides some descriptive information on the six measures of usage, for the in- and out-of-sample data together. The measure present a diversity of characteristics - three of the measures present a high percentage of zeros, but three have more moderate percentages of zeros. There is considerable variation in the means of the measures, and in their maximums. Looking at the mean/variance ratio se see that all of the variables are unconditionally overdispersed, some strongly so, and some relatively moderately so.

4 Results

Tables II through VII contain the results for the six measures of health care use. In these tables, "Type" refers to the use of a NB-I or NB-II model, for the NB, HNB, MNB and CMNB models. For the models that use a series expansion, "Order" refers to the degree of the shaping polynomial, *e.g.* the p in equation 5. To save space, we report only the results for the best fitting type and order for each model, according to the values of the consistent Akaike (CAIC), Bayesian (BIC), and the Akaike (AIC) information criteria. The information criteria results are all divided by the lowest value across models of the respective criterion, to facilitate comparisons. We also report the in and out-of-sample mean squared prediction error (MSE), relative to that of the best fitting model. Since the best order may vary according to the different information criteria,

the order is reported for each of the information criteria. The best type may also differ for in- and out-of-sample relative MSE, so it is reported. The best model, according to each of the information criteria and the two relative MSE measures is indicated by bold highlighting. For example, in Table II, the best model according to the CAIC is the NBSNP using a NB-II baseline model, with order $p = 2$. The best model according to the AIC is the HPSP model with order 1 for the binary part of the model and order 2 for the zero-truncated count part of the model. The HNB-II model has minimum relative MSE, both in- and out-of-sample. In the second column, the 2;2;2 in the line for the PSP model means that $p = 2$ was best according to the CAIC, BIC, and AIC measures, respectively. The (1,2);(1,2);(1,2) in the line for the HPSP model gives the optimal order as a pair, first for the binary part of the model, then for the zero-truncated part of the model.

Comparing Tables II through VII, the most striking general conclusion is that the models, with the exception of the PSNP model in the case of several measures of use, achieve very similar values for both the information criteria and for relative MSE. Differences of more than two percent for the CAIC and BIC are fairly uncommon, and the differences are even smaller for the AIC. In many cases the differences are so small as to be almost surely statistically insignificant. Bootstrapping confirms that, for the OBDV use measure, the CAIC, BIC and AIC measures for the NB-I and NBSNP-II(2) models are not significantly different at the 90% confidence level. Furthermore, the 90% bootstrap confidence interval for CAIC, BIC and AIC for the NB model contains the respective CAIC, BIC and AIC values for all models and all six usage measures, with the exception of the PPP model in the case of the OBDV, DV and PRESCR usage measures. [11]

Across the six measures of use, the MNB, CMNB, HPSP and PSNP models

[11] Bootstrapping is not feasible on a large scale due to the overwhelming computational burden it would entail. This is because some of the models, such as the MNB and HPSP, can become numerically unstable, and others, such as the PSNP and NBSNP, can present multiple local optima.

are never preferred according to the CAIC and BIC criteria, though the CMNB and HPSP models are preferred for one measure each according to the inconsistent AIC criterion. Looking at relative MSE, the differences are almost always so small that it does not seem warranted to discriminate between models on this basis. Remarkably, the PSNP model performs best for out-of-sample fit in the case of the PRESCR measure of use, in spite of the fact that it is strongly dominated in terms of the information criteria.

Other results of note include the fact that the CMNB model outperforms the more heavily parameterized MNB model in almost all cases. One may also note that NB-I and NB-II models, on their own or as the basis for the MNB, HNB and NBSNP models, are preferred a roughly equal number of times. Also, superiority of a Type I or Type II model in terms of the information criteria is at best loosely linked to superiority in terms of out-of-sample MSE. Another observation is that the optimal order of the semiparametric and semi-nonparametric models that are based on the NB model is almost always $p = 1$ or $p = 2$, according to the two consistent information criteria.

Having seen that the models perform quite similarly in terms of information criteria and fit, we ask whether the choice of the model, among the models under consideration, will importantly alter the conclusions of a policy analysis.[12] To offer a partial answer to this question, first we examine the impact of income on usage of health care services, unconditional on other explanatory variables. This is done by grouping fitted usage by income levels, then averaging the fitted value for each group to marginalize out the other explanatory variables.[13] This is done using all the observations, both in- and out-of-sample.

In Figures 1 through 6 we see that the models predict levels that are quite

[12] Cameron and Johansson (1997) offer limited evidence on this issue in their Tables IV and VI, where they compare average derivatives for several statistical models. They find that the average derivatives are quite similar for most conditioning variables.

[13] It would be possible to present plots of fitted usage versus income, or any other variable, holding the other variables fixed at given levels. A large (infinite) number of such plots could be generated, which is the reason we use the simple approach of marginalizing the other variables.

similar across the models, especially for the ERV, IPV and DV measures.[14] Also, the discrete analog of a derivative almost always has the same sign across models, for all measures. The broad conclusions one draws do not seem to be model-dependent. To comment on the results, one may note that ERV and IPV generally decline with income, which may reflect the effects of poverty on the predisposition to use emergency means of care rather than preventive methods. Dental visits (DV) and office-based visits (OBDV) both generally increase with income, which is probably due to the more discretionary nature of such usage. One may note that PRESCR and OBDV share a similar shape, which is not unexpected given that a doctor will usually prescribe medication following a visit. These conclusions hold for all of the models under consideration.

We repeat this exercise by plotting predicted usage versus age categories in Figures 7 through 12. We again see that the broad picture is the same independent of the model. The measures of use, with the exception of emergency room visits, which presents a U-shaped form, rise steadily with age.

5 Conclusions

Most of the statistical models considered in this paper are considerably more flexible than the simple Poisson and negative binomial models which have been the standards for count data for many years. Nevertheless, the more flexible models do not significantly outperform the negative binomial model in terms of information criteria or out-of-sample prediction error. As well, the negative binomial model leads to predicted visits conditional on income and age that are very similar to those of the other models. This is true for all six of the dependent variables we considered, which, one will recall, encompass a variety of characteristics. The models as a group perform quite similarly according to these criteria.

These facts suggest that applied researchers conducting a policy analysis

[14] One should note that the abscissa of some of the Figures does not begin at zero.

may not need to be excessively preoccupied with which among these models they base their analysis. The models will lead to similar results in terms of their predictions, at least for data similar to that we employ. It seems that effort will be better spent on the choice of the conditioning variables to include, how these variables are to be measured, and how the model is to be parameterized as a function of these variables. Issues of endogeneity should not be overlooked, either, though we have abstracted from this.

These statements are not without a caveat. The models *do* lead in some cases to substantially different estimates of conditional and unconditional probabilities that a given number of visits are taken, with the important differences usually being the probabilities of zeros and low numbers of visits. The negative binomial model often underpredicts zeros, as is well known. Cost of care may depend on the number of individuals treated as well on the total number of visits they make, since files must be opened and records kept for each individual. If this is an important feature of the analysis, then an accurate prediction of the number of users of a service among a population will be important. This will require an accurate predicition of the number of zero visits. In this case, the choice of the model will be more important. However, among the more flexible models there are only minor differences in the way they fit specific count probabilities, so amongst this group the choice of model will not be very important even when the goal is to fit count probabilities.

From the point of view of theoretical work, it may be more worthwhile to try to find solutions to problems of endogeneity of regressors in count models that lead to more precise estimates than those that result from GMM applied to low order moments than to further develop flexible densities to be using in univariate MLE. It seems that the currently available models do a good job of extracting the information the data contains.

References

[1] Cameron, A.C. and P.K. Trivedi (1986), Econometric models based on count data: comparisons and applications of some estimators and tests, *Journal of Applied Econometrics*, **1**, 29-54.

[2] Cameron, A.C. *et. al.* (1988), A microeconometric model of the demand for health insurance and health care in Australia, *Review of Economic Studies*, **55**, 85-106.

[3] Cameron, A.C. and P. Johansson (1997), Count data regression using series expansions: with applications, *Journal of Applied Econometrics,* **12**, 203-23.

[4] Cameron, A.C. and P. Johansson (1998), Bivariate count data regression using series expansions: with applications, mimeo, http://www.econ.ucdavis.edu/faculty/cameron/research/bv3a.pdf

[5] Cragg, J.G. (1971), Some statistical models for limited dependent variables with applications to the demand for durable goods, *Econometrics*, **39**, 829-44.

[6] Deb, P. and P.K. Trivedi (1997), Demand for medical care by the elderly: a finite mixture approach, *Journal of Applied Econometrics,* **12**, 313-36.

[7] Doornik, J.A. (1999), Object-oriented matrix programming using Ox, 3rd ed. London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

[8] Geil, P. *et. al.* (1997), Economic incentives and hospitalization in Germany, *Journal of Applied Econometrics,* **12**, 295-311.

[9] Gallant, A.R. and D.W. Nychka (1987), Seminonparametric maximum likelihood estimation, *Econometrica*, **55**, 363-90.

[10] Gerdtham, U-G., Equity in health care utilisation: further evidence based on hurdle models and Swedish macro data, *Health Economics*, **6**, 303-19.

[11] Gerdtham, U-G. and P.K. Trivedi (2000), Equity in Swedish health care reconsidered: new results based on the finite mixture model, mimeo, http://swopec.hhs.se/hastef/abs/hastef0365.htm

[12] Grossman, M. (1972), On the concept of health capital and the demand for health, *Journal of Political Economy,* **80**, 223-55.

[13] Gurmu, S. (1997), Semi-parametric estimation of hurdle regression models with an application to medicaid utilization, *Journal of Applied Econometrics,* **12**, 225-42.

[14] Gurmu, S. and P.K. Trivedi (1996), Excess zeros in count models for recreational trips, *Journal of Business and Economic Statistics*, **14**, 469-77.

[15] Gurmu, S. *et. al.* (1999), Semiparametric estimation of count regression models, *Journal of Econometrics*, **88**, 123-50.

[16] Manning, W.G. *et. al.* (1987), Health insurance and the demand for medical care: evidence from a randomized experiment, *American Economic Review*, **77**, 251-77.

[17] Mullahy, J. (1986), Specification and testing of some modified count data models, *Journal of Econometrics*, **33**, 341-65.

[18] Pohlmeier, W. and V. Ulrich (1995), An econometric model of the two-part decision-making process in the demand for medical care, *Journal of Human Resources*, **30**, 339-61.

[19] Romeu, A. and A.M. Vera-Hernández, A semi-nonparametric estimator for counts with an endogenous dummy variable, mimeo, http://pareto.uab.es/wp/2000/45200.pdf

[20] Vera-Hernández, A.M. (1999), Duplicate coverage and demand for health care. The case of Catalonia, *Health Economics*, **8**, 579-98.

[21] Windmeijer, F.A.G. and J.M.C. Santos Silva (1997), Endogeneity in count data models: an application to demand for health care, *Journal of Applied Econometrics,* **12**, 281-94.

[22] Zweifel, P. (1981), Demande medicale induite par l'offre: chimere ou realite (Supplier induced demand in a model of physician behaviour), *Consommation* **28**, 39-62.

Table I: Descriptive Statistics

	OBDV	OPV	ERV	IPV	DV	PRESCR
Observations	4562	4565	4566	4566	4566	4486
Mean	3.189	0.243	0.194	0.086	1.053	6.476
St. Dev.	5.343	1.589	0.637	0.389	1.875	11.010
Mean/Variance	0.112	0.096	0.478	0.568	0.299	0.053
Maximum	68	50	17	5	32	75
% Zeros	34.6	90.4	86.1	93.9	57.1	33.8

Table II: Information Criteria and Fit, OBDV

Model	Type/Order	CAIC	BIC	AIC	Type/Order	INSAMP	OUTSAMP
NB	I	1.0034	1.0034	1.0086	I;I	1.007	1.005
HNB	I	1.0030	1.0023	1.0032	II;II	**1.000**	**1.000**
MNB	I	1.0037	1.0028	1.0025	I;I	1.005	1.010
CMNB	I	1.0005	1.0002	1.0032	I:I	1.008	1.007
PSP	2;2;2	1.0025	1.0025	1.0063	1;1	1.007	1.011
HPSP	(1,2);(1,2);(1,2)	1.0025	1.0014	**1.0000**	(1,1);(1,1)	1.000	1.000
PSNP	4;4;4	1.1928	1.1928	1.1979	2;2	1.002	1.003
NBSNP	II, 2;2;2	**1.0000**	**1.0000**	1.0037	(II,1);(II,4)	1.005	1.005

Table III: Information Criteria and Fit, OPV

Model	Type/Order	CAIC	BIC	AIC	Type/Order	INSAMP	OUTSAMP
NB	II	1.0079	1.0091	1.0128	II;II	1.004	1.004
HNB	II	1.0260	1.0227	1.0093	II;II	1.003	1.005
MNB	I	1.0260	1.0227	1.0035	I;I	1.004	1.005
CMNB	II	1.0045	1.0034	1.0012	II;II	1.006	1.003
PSP	1;1;2	**1.0000**	**1.0000**	**1.0000**	1;3	1.004	**1.000**
HPSP	(1,2);(1,2);(1,2)	1.0339	1.0295	1.0035	(1,1);(1,1)	1.005	1.006
PSNP	3;3;3	1.0959	1.0952	1.0957	4;1	**1.000**	1.004
NBSNP	I, 2;2;2	1.0034	1.0023	1.0023	(II,2);(II,1)	1.002	1.004

Table IV: Information Criteria and Fit, ERV

Model	Type/Order	CAIC	BIC	AIC	Type/Order	INSAMP	OUTSAMP
NB	I	**1.0000**	**1.0000**	1.0020	I;II	1.000	1.000
HNB	I	1.0212	1.0183	1.0020	II;II	1.001	**1.000**
MNB	I	1.0327	1.0290	1.0089	I;I	1.000	1.000
CMNB	I	1.0087	1.0068	1.0010	I;II	1.000	1.000
PSP	1;1;2	1.0048	1.0039	1.0020	3;1	1.000	1.000
HPSP	NC	NC	NC	NC	NC	NC	NC
PSNP	2;2;4	1.0067	1.0058	1.0030	1;1	1.000	1.000
NBSNP	I,1;1;2	1.0038	1.0029	**1.0000**	I;II	**1.000**	1.000

Table V: Information Criteria and Fit, IPV

Model	Type/Order	CAIC	BIC	AIC	Type/Order	INSAMP	OUTSAMP
NB	II	**1.0000**	**1.0000**	1.0019	I;I	1.000	1.000
HNB	II	1.0426	1.0354	1.0058	II;II	1.002	1.010
MNB	I	1.0519	1.0447	1.0058	I;I	1.025	1.033
CMNB	I	1.0167	1.0130	**1.0000**	I;II	1.002	1.007
PSP	1;1;1	1.0074	1.0056	1.0019	1;1	1.002	1.006
HPSP	(1,1);(1,1);(1,1)	1.0611	1.0503	1.0078	(1,1),(1,3)	1.005	1.005
PSNP	2;2;2	1.0056	1.0037	1.0019	1;1	**1.000**	1.007
NBSNP	II, 1;1;3	1.0056	1.0056	1.0019	(I,1);(I,3)	1.000	**1.000**

Table VI: Information Criteria and Fit, DV

Model	Type/Order	CAIC	BIC	AIC	Type/Order	INSAMP	OUTSAMP
NB	I	1.0019	1.0023	1.0065	II;II	1.003	1.004
HNB	I	1.0045	1.0042	1.0019	I;I	**1.000**	1.001
MNB	I	1.0117	1.0110	1.0065	I;I	1.000	1.007
CMNB	I	1.0053	1.0053	1.0069	I;I	1.004	1.007
PSP	2;2;3	1.0189	1.0193	1.0218	1;1	1.003	1.004
HPSP	(1,2);(1,2);(1,4)	1.0076	1.0064	**1.0000**	(1,4);(1,4)	1.001	**1.000**
PSNP	2;4;4	1.0615	1.0310	1.0325	1;1	1.002	1.003
NBSNP	I, 3;3;3	**1.0000**	**1.0000**	1.0015	(II,1);(II,1)	1.002	1.004

Table VII: Information Criteria and Fit, PRESCR

Model	Type/Order	CAIC	BIC	AIC	Type/Order	INSAMP	OUTSAMP
NB	I	1.0013	1.0021	1.0053	II;II	1.005	1.003
HNB	II	**1.0000**	**1.0000**	**1.0000**	II;II	**1.000**	1.001
MNB	I	1.0013	1.0011	1.0002	I;I	1.010	1.009
CMNB	I	1.0011	1.0017	1.0034	II;II	1.009	1.006
PSP	1;1;2	1.0130	1.0136	1.0161	1;1	1.038	1.035
HPSP	(1,1);(1,1);(1,1)	1.0068	1.0066	1.0051	1;1	1.024	1.023
PSNP	2;2;2	1.8564	1.8581	1.8670	1;2	1.004	**1.000**
NBSNP	I, 2;2;2	1.0006	1.0009	1.0034	(II,1);(II,2)	1.005	1.003

Figure 1: Fitted Use Versus Income: OBDV

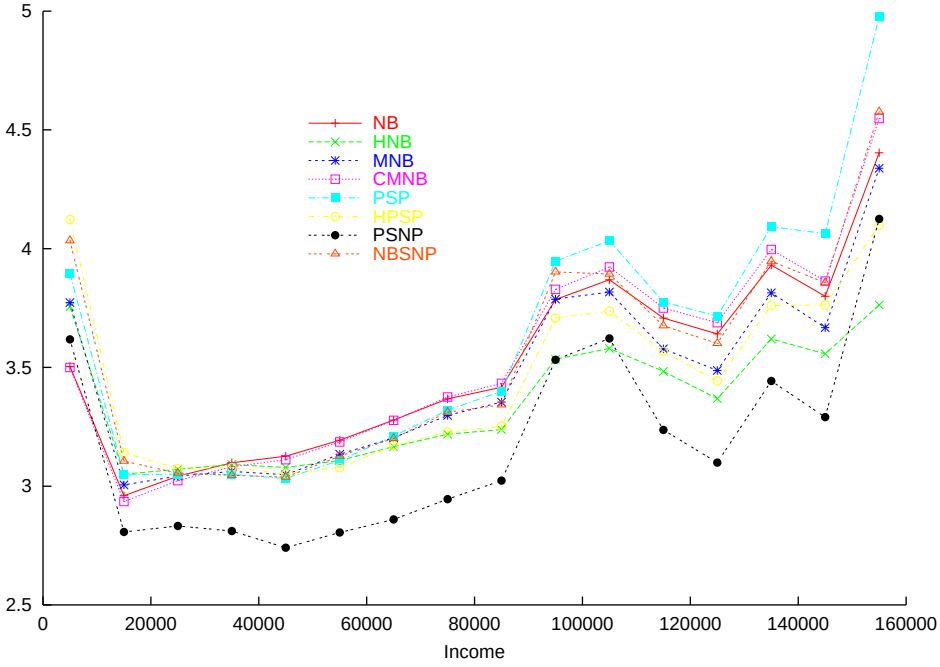

Figure 2: Fitted Use Versus Income: OPV

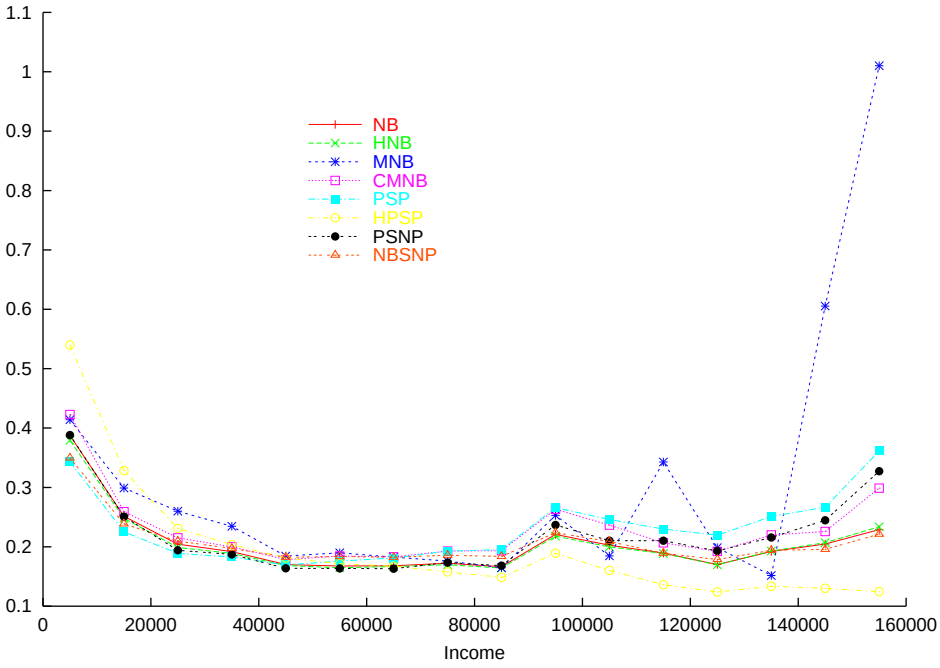

Figure 3: Fitted Use Versus Income: ERV

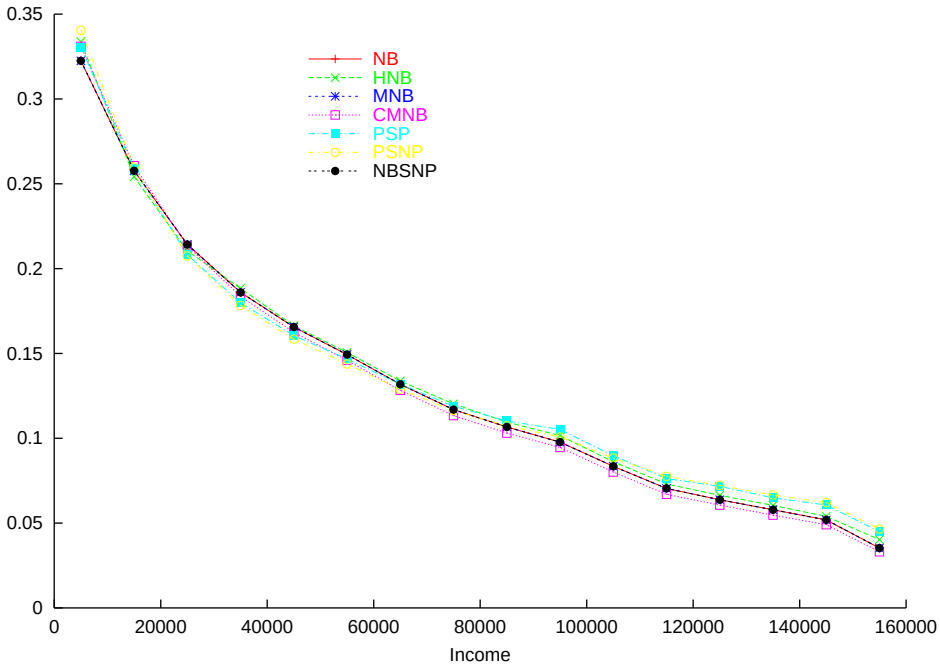

Figure 4: Fitted Use Versus Income: IPV

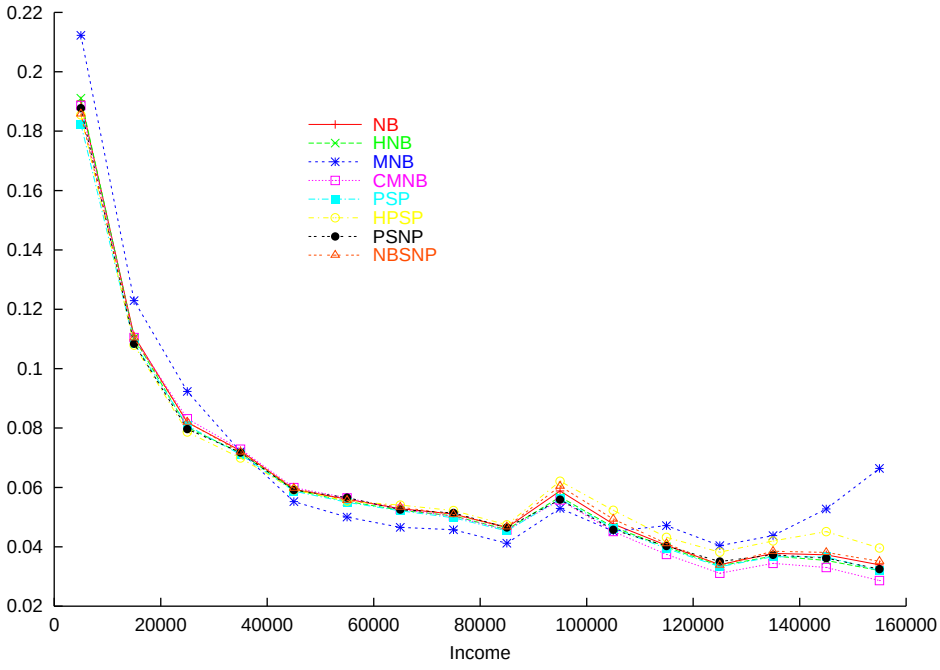

Figure 5: Fitted Use Versus Income: DV

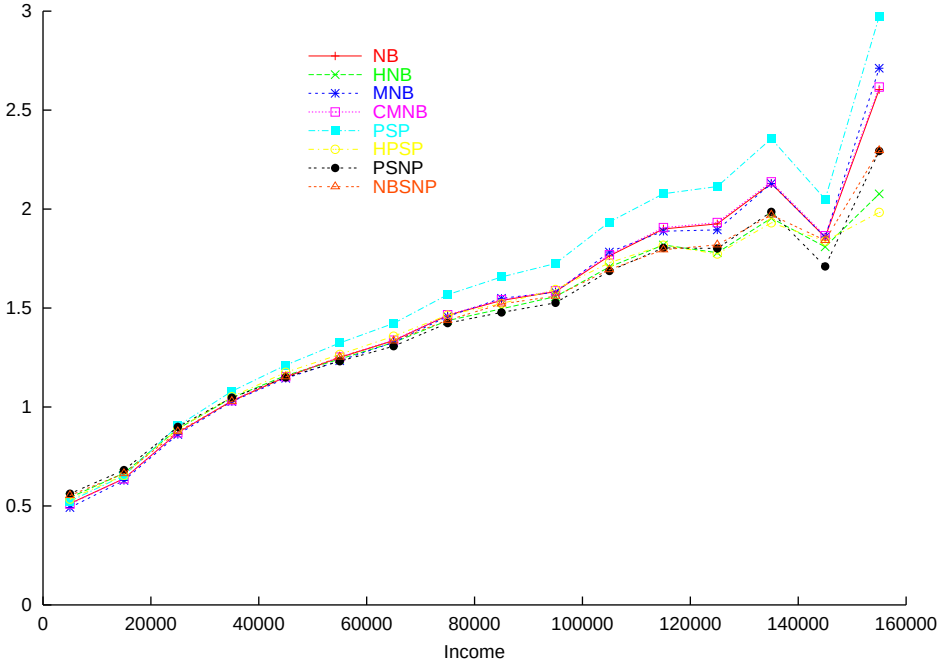

Figure 6: Fitted Use Versus Income: PRESCR

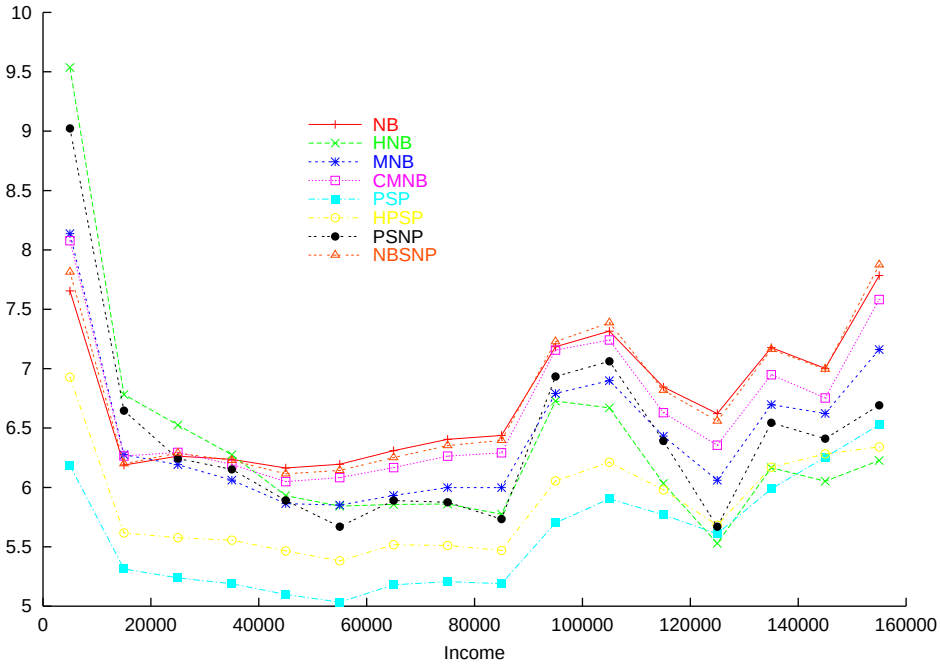

Figure 7: Fitted Use Versus Age: OBDV

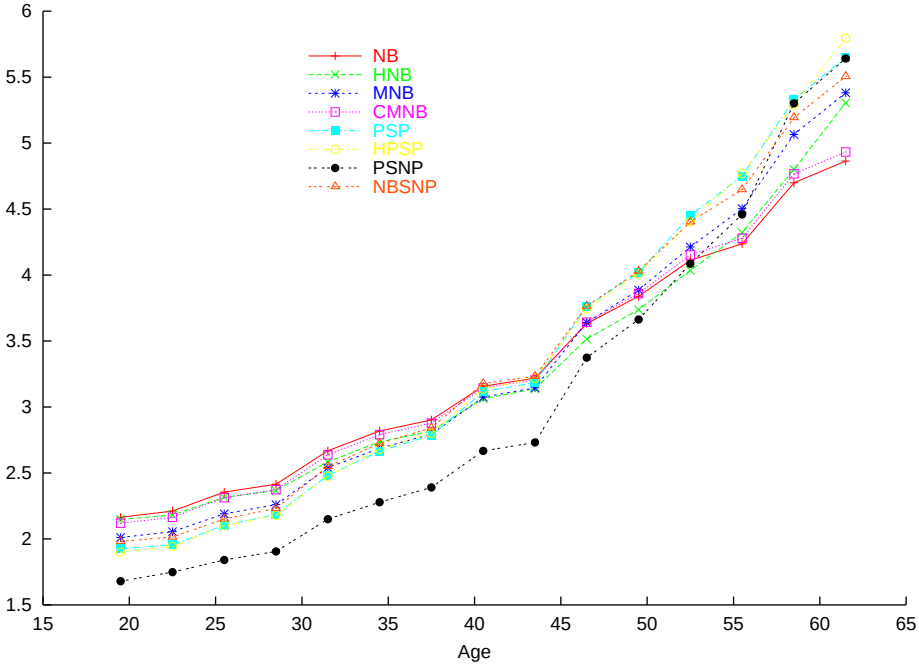

Figure 8: Fitted Use Versus Age: OPV

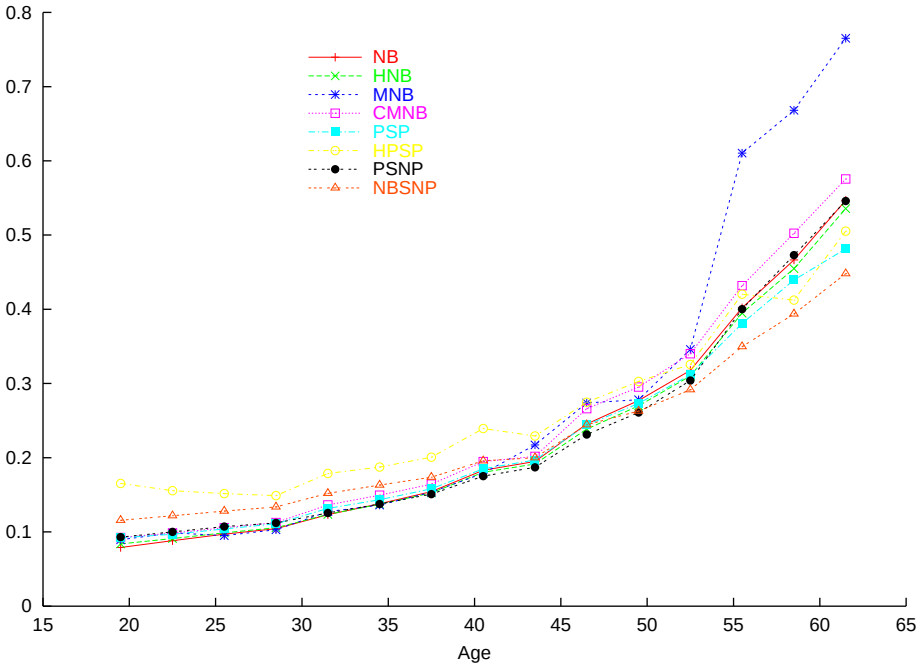

Figure 9: Fitted Use Versus Age: ERV

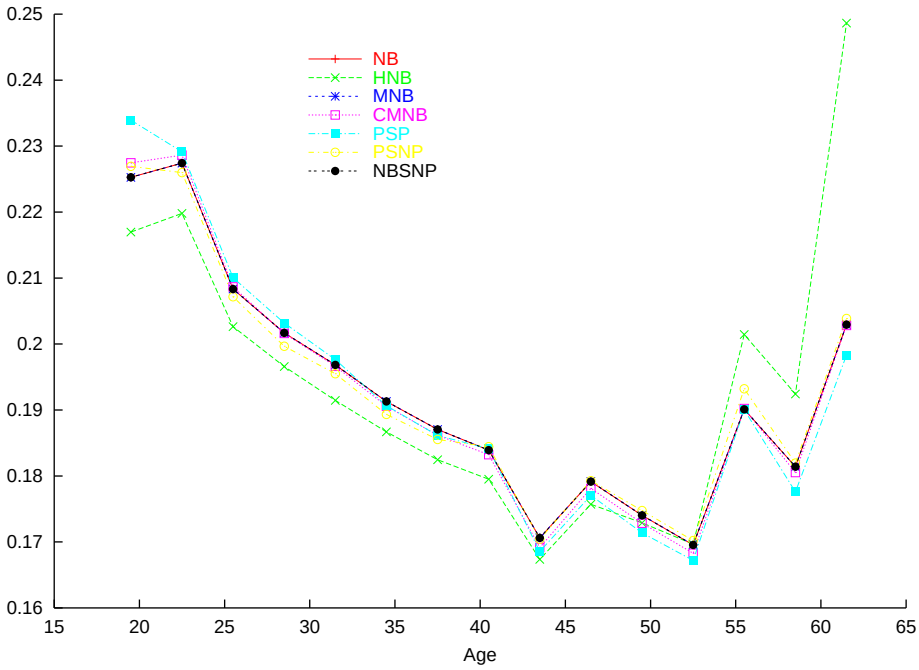

Figure 10: Fitted Use Versus Age: IPV

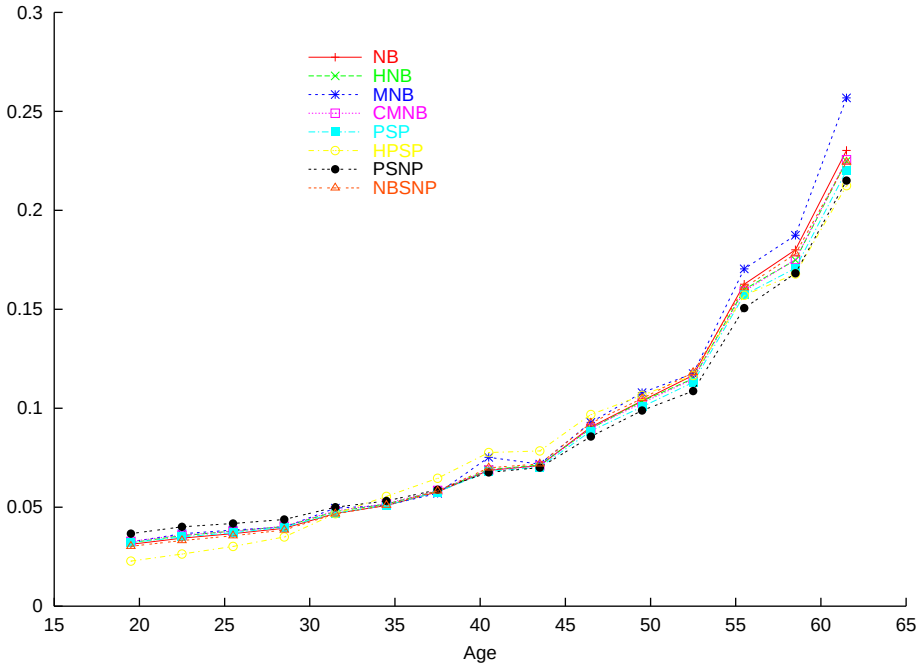

Figure 11: Fitted Use Versus Age: DV

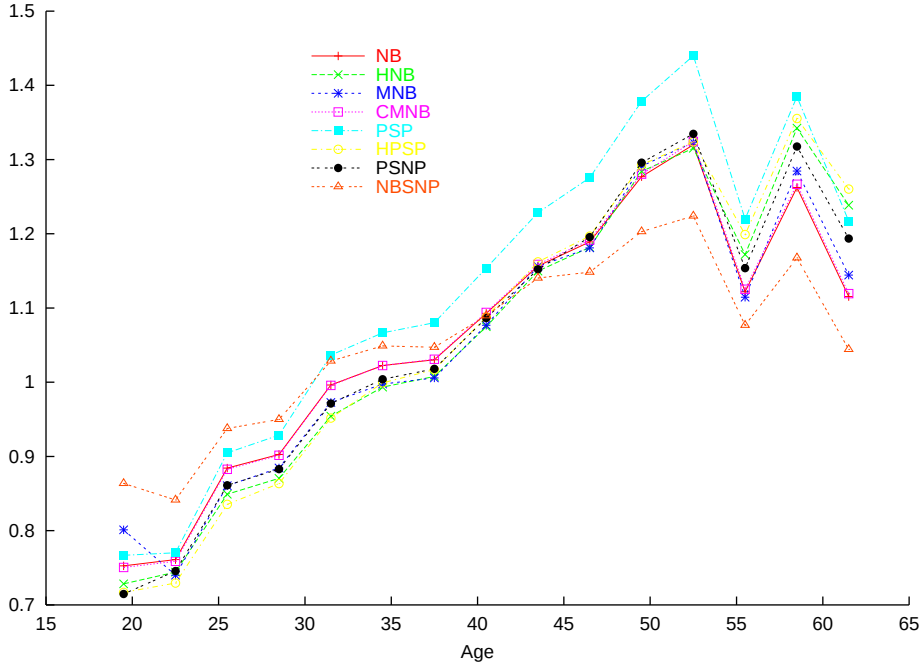

Figure 12: Fitted Use Versus Age: PRESCR